VOLUME 39 JUNE 1953 NUMBER 6

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
OF THE
UNITED STATES OF AMERICA

EDITORIAL BOARD

LINUS PAULING, *Chairman* EDWIN B. WILSON, *Managing Editor*
ALEXANDER WETMORE, *Home Secretary* ROGER ADAMS, *Foreign Secretary*
WILLIAM W. RUBEY, *Chairman of the National Research Council*

J. W. BEAMS
R. E. CLELAND
W. W. RUBEY
HARLOW SHAPLEY
F. E. TERMAN
G. W. BEADLE
E. A. DOISY
G. C. EVANS
C. H. GRAHAM
J. T. PATTERSON
GREGORY BREIT
L. C. DUNN
W. F. LIBBY
J. VON NEUMANN
PAUL WEISS

Publication Office: Mack Printing Company, Easton, Pa.
Editorial Office and Home Office of the Academy
2101 Constitution Avenue, Washington 25, D. C.

Entered as Second-Class matter, June 21, 1920, at the Post-office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in section 1103, act of October 3, 1917. Authorized June 18, 1920.

Published Monthly Price per Annum, $7.50

OFFICERS OF THE ACADEMY

PRESIDENT ..... DETLEV W. BRONK
VICE PRESIDENT ..... EDWIN B. WILSON
FOREIGN SECRETARY ..... ROGER ADAMS
HOME SECRETARY ..... ALEXANDER WETMORE
TREASURER ..... WILLIAM J. ROBBINS

INFORMATION TO SUBSCRIBERS

Checks covering subscriptions to the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES at the rate of $7.50 per annum should be made payable to the National Academy of Sciences, and sent to the Treasurer, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C. Single numbers $0.75. These prices effective January, 1951.

Microfilm copies of the PROCEEDINGS are available to subscribers to the paper edition. The microfilm is a positive, furnished on metal reels, suitably labeled, and is distributed at the end of the volume year. Inquiries concerning purchase should be directed to University Microfilms, 313 N. First St., Ann Arbor, Michigan.

Past publications of the NATIONAL ACADEMY OF SCIENCES from 1863 to 1926 and of the NATIONAL RESEARCH COUNCIL from 1916 to 1925 are listed in the Index to the First Ten Volumes of these PROCEEDINGS. This Index appeared as Part 2 of Volume 13, No. 1, January, 1927, and is for sale separately at the price of $1.50. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

Past volumes of the PROCEEDINGS may be obtained at five dollars per volume unless the sale of the volume would break a complete set of Volumes I to XXX. Single numbers may be obtained for fifty cents except where the sale of such numbers would break up a complete volume. Beginning with Volume 37 (1951), complete volumes are $7.50, and single numbers $0.75. Only forty complete sets of the PROCEEDINGS are available for sale —Volumes I to XXX price $200.00. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

THE NATIONAL ACADEMY OF SCIENCES-NATIONAL RESEARCH COUNCIL also issues technical publications reviewing and correlating recent research. The *Bulletin* and *Reprint and Circular Series* have been replaced by a single publication series, available individually, or by subscription at 25% discount. For a list of available publications, with prices and ordering procedures, address the Publications Office, National Academy of Sciences-National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 39 June 15, 1953 Number 6

SEROLOGICAL REACTIVITY OF SYNTHETIC POLYGLUCOSES

By Michael Heidelberger and Alan C. Aisenberg

Dept. of Medicine, Columbia University College of Physicians and Surgeons, and the Presbyterian Hospital, New York

Communicated March 9, 1953

Because of the great complexity of natural substances with antigenic properties little is known of the chemical basis of their immunological specificity. Methods for the determination of the fine structure of the proteins, the class of substances to which many antigens belong, are still in their infancy. Knowledge of the particularities of the structure of carbohydrates is, however, in a more advanced state. Within this class of substances certain members, such as the type-specific capsular polysaccharides of pneumococci,[1] not only interact with antibodies to yield precipitates, but may even stimulate the production of antibodies in, for example, man and the mouse. The constituent sugars of several of the type-specific polysaccharides are known, and the fine structure of the specific polysaccharide of Type III pneumococcus has been worked out,[2] as has also that of one of the specific carbohydrates of the tubercle bacillus.[3] Moreover, the development of quantitative, absolute methods for the estimation of antibodies[4] provided a new tool for the study of immune reactions and led to a quantitative theory of the reaction of specific precipitation,[5] in terms of which the cross-reactivities of chemically related specific polysaccharides such as those of pneumococcus Types III and VIII could be interpreted.[6]

The recent production of synthetic polyglucoses with possible uses as plasma extenders furnished the stimulus for the studies which follow.* Dextran, a series of polyglucoses built up by various strains of *Leuconostoc*, is known to give precipitates in antipneumococcus sera of Types II, XII, XX and, to a smaller extent, XXII.[7] Severe reactions occasionally encountered in the use of dextran in human beings have been ascribed to its possible reactivity with preexisting antibodies.[8] A serological study of the synthetic polyglucoses was therefore of practical as well as theoretical interest.

Experimental.—A polymer of glucose, 52R611, kindly supplied by Merck & Co., Inc., Rahway, N. J.,† was fractionated as follows: 10 g. were added to 100 ml. of 5% sodium acetate solution. The dark mixture was brought to pH 6.5 by addition of sodium hydroxide and allowed to stand in the cold for two days, with occasional mixing. All subsequent operations were also carried out in the cold. The soluble portion, at a volume of 76 ml., was mixed with 49 ml. of chilled isopropyl alcohol[9] which sufficed to form a precipitate. After two days this was centrifuged off and taken up in water. The centrifuged solution and the washings of a residual gel were poured into several volumes of redistilled ethyl alcohol and the precipitate was filtered off and dried in vacuo (fraction A). The original supernatant from which A had separated was treated with 20 ml. more of isopropyl alcohol and the precipitate (B) isolated as was A. The supernatant from B was made up to about 225 ml. with 95 per cent ethyl alcohol and the precipitated fraction (D) was isolated as before. Fraction A was divided roughly

TABLE 1

FRACTIONATION OF 10 G. OF GLUCOSE POLYMER 52R611

FRACTION	YIELD, G.	$\eta_{rel}^{25°a}$	$\eta_{red}^{25°}$	REDUCING SUGAR AS GLUCOSE, UNHYDROLYZED, %	MAXIMUM ANTIBODY N, %, PPTD. FROM ANTISERUM, TYPE			
					IX	XII	XX	XXII
A	0.19	1.182	3.0	4.6	35	50	50	35
A′	0.14	1.075	1.25	5.0	21	35	35	25
B	1.16	1.061	1.0	4.3	19	30	25	20
D	1.45	1.030	0.5	4.3	3	15	6	4
E	1.8[b]	1.027	0.45	4.0	0.1	5	3	0.2

[a] 0.6% solution in 0.85% NaCl solution. "rel" and "red" refer to relative and reduced viscosities, respectively.

[b] Also per 10 g. polyglucose.

in half by partial precipitation from 10 ml. of 5 per cent sodium acetate solution with glacial acetic acid, a method known to be useful in the separation of serologically active from inactive polysaccharides.[10] The insoluble fraction was designated A and the soluble portion A′. Properties of the four fractions are summarized in table 1, and their quantitative behavior toward a Type XII antipneumococcus horse serum is shown in figure 1. Plots of data obtained in antisera to Types IX, XX, and XXII were remarkably similar, and are therefore not given.

In an attempt to find a fraction without serological reactivity 20 g. of polymer 52R611 were treated as above for the separation of fraction A. Material corresponding to fraction B and part of D was then precipitated in the cold by bringing the isopropyl alcohol concentration to 51 per cent by volume instead of to 47.5 per cent, as in the first instance. All residual polymer precipitable by alcohol (E) was then thrown down by bringing the volume to 1 liter with 95 per cent ethanol in the cold. The yield was 3.6 g.

TABLE 2

CROSS REACTIONS OF SYNTHETIC POLYGLUCOSES AND DEXTRANS IN ANTIPNEUMOCOCCUS HORSE SERA

ANTISERUM TYPE →	I	II	III	IV	V	VI	VII	IX	X	XI	XII	XIV	XV	XVIII	XIX	XX	XXII
Synthetic polyglucose																	
52R2109[a]	..	+++	+±	..	..	−	+++	+++	+±	+±	+++	++	+±	++	++	+++±	++++
52R2108[b]	+±	+++±	..	+±	−	−	+++	..	..	+±	++++	−	..	+++±	−	+++	++++
52R5413	..	±(+)[c]	..	..	..	..	±(±)	+	..	−(+)	++	..	..	+	..	++	+±
52R5414	..	−(−)	..	−(−)	..	..	−(+)	−(+)	..	−(+)	+±	..	..	+	..	+±	+
52R5416	..	++	..	..	..	..	+±	++	..	+	+±	..	..	+	..	+++	++±
52R5418	..	±(+)	..	..	..	..	±(+)	+±	..	+	+±	..	..	±	..	++	++
P218–56[d]	..	−	..	..	..	..	−	−	..	−	+++	..	..	−	..	+±	−
Dextran																	
B742	..	++	..	..	..	..	+±	++±	..	+	++	..	..	++	..	++±	+++
N236	..	+++	..	..	..	..	+±	+±	..	−	+++	..	..	++±	..	++±	++±
N150N	..	++±	..	..	..	..	±	−	..	−	+	..	..	+±	..	++	++±

[a] Reduced viscosity, 0.33.
[b] Reduced viscosity, 0.59.
[c] Values in parentheses indicate readings after centrifugation.
[d] Mol. wt., number average, 21,000–46,000.

As noted in tables 1 and 3 and figure 1, fraction E gave considerably less precipitate with antisera than did the others.

In table 2 are shown qualitative tests of other samples from Merck and duPont,[‡] some of them fractionated, as 52R2108, 9, others partially hydrolyzed and/or prepared with different catalysts. Table 3 contains data on the total antibody (antibody to the group-specific "C" polysaccharide as well as type-specific antibody) precipitated by many of the fractions. Antibody determinations were carried out mainly at 0°C. according to

TABLE 3

QUANTITATIVE DATA ON HOMOLOGOUS AND CROSS-REACTING ANTIBODIES IN VARIOUS ANTIPNEUMOCOCCUS HORSE SERA

ANTISERUM, TYPE	ANTIBODY N, UNABSORBED SERUM, μG./ML.	ANTIBODY N TO Pn C-SUBSTANCE, μG./ML.	ANTIBODY N IN C-ABSORBED SERUM, μG./ML.	ANTIBODY N PPTD. BY 52R2109 FROM: UNABSORBED, μG./ML.	ANTIBODY N PPTD. BY 52R2109 FROM: C-ABSORBED SERUM, μG./ML.	ANTIBODY N PPTD. BY 52R2108 FROM: UNABSORBED, μG./ML.	ANTIBODY N PPTD. BY 52R2108 FROM: C-ADSORBED SERUM, μG./ML.	ANTIBODY NITROGEN PPTD BY 52R611: FRACTION A FROM UNABSORBED SERUM,[c] μG./ML.	FRACTION B FROM UNABSORBED SERUM, μG./ML.	FRACTION D FROM UNABSORBED SERUM, μG./ML.	FRACTION E FROM UNABSORBED SERUM, μG./ML.	ANTIBODY N PPTD. BY P218-56 FROM C-ABSORBED SERUM, μG./ML.
I	...	70	1000	9	8	12	...	...	...	...	...	...
II	...	98	1061	50	28	66	56	...	...	...	...	...
VII	1030[a]	580	< 428	64	25	...	...	< 156	...	...	...	...
IX	1655	32	...	< 285	...	< 350	...	570[d]	310	48	2	0
XI	794	36	...	...	9	...	...	...	...	...	...	...
XII (625)	1780, 1853[b]	350, 415	1360	429	< 345	< 596	...	900[e]	550	262	88	138
XII (296)	1544	189	...	...	...	...	...	883[f]	...	...	...	...
XVIII	2218	11	...	...	20	...	...	...	...	...	...	...
XX	468	27	...	108	< 91	...	...	240	112	28	13	12
XXII	847	14	...	...	122	...	...	280	170	30	2	1

[a] 850 with specific polysaccharide of pneumococcus Type VII (S VII) treated with HNO_2 to remove C-substance. Supernatant gave 365 μg. with C; this supernatant gave 7 μg. N with A.

[b] 1764 with specific polysaccharide of pneumococcus Type XII (S XII) treated with HNO_2 to remove C-substance. Supernatant gave 149 μg. with C; this supernatant gave no ppt. with A.

[c] Fraction A' gave values slightly higher than those with Fraction B.

[d] The supernatant gave 1143 μg. antibody N with S IX; total, 1713.

[e] The supernatant gave 950 μg. antibody N with S XII. 20 mg. Fraction D per ml. antiserum, in the antigen-excess zone of complete inhibition, gave rise to 85% inhibition of precipitation by 2 mg. Fraction A; precipitation by S XII was not affected.

[f] 776 μg. N were precipitated from the C-absorbed serum by Fraction A.

reference 11 or, with smaller quantities, with estimation of N by the Markham method.[12] A few analyses were also run at 37°C. for comparison (Fig. 1), since cross reactions are normally more influenced by differences in temperature than are the homologous reactions.[6]

Discussion.—It is evident that the synthetic polyglucoses give precipitates in some, but not all, of a wide range of antipneumococcus horse sera. That it is actually the antibody in these sera which is precipitated is shown in several ways: first, by the smaller amount of nitrogen thrown down

after removal of antibodies to the group-specific "C" substance of pneumococcus, and second, by more detailed analyses, mainly with Type XII serum. These showed that after removal of the antibodies with the specific polysaccharide of Type XII pneumococcus and with pneumococcal "C" substance the residual serum gave only negligible precipitation with the otherwise maximally reactive polyglucose fraction A. Also, in the antisera to several types, the amount of antibody directly precipitable by the pneumococcal polysaccharides usually agreed well with the sum of the quantities precipitated first by a polyglucose and then by the pneumococcal carbohydrates (cf. footnotes to table 3).

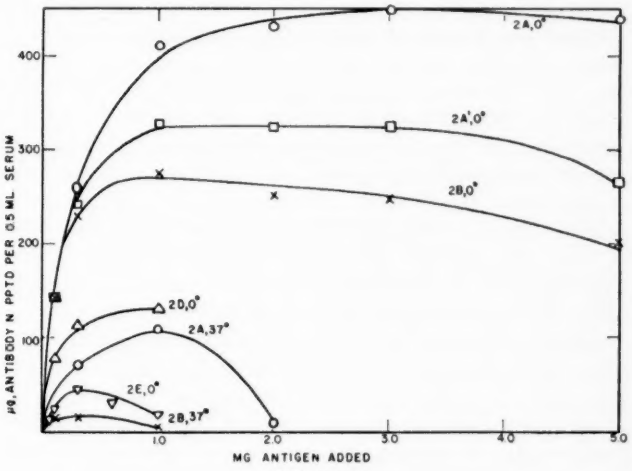


FIGURE 1

Reactions of polyglucose fractions A, A′, B, D, E with antipneumococcus Type XII Horse Serum, No. 625.

The reducing values of the fractions given in table 1 indicate a similar proportion of such end groups, although the viscosities point to large differences in size or structure, or both. In view of the rather random distribution of mainly 1,4- and 1,6-linkages found by Pacsu and his coworkers[13] in unfractionated material with molecular weight averaging 180,000 (by the method of light scattering) it is possible, though not certain, that the products of highest viscosity represent those of largest molecular size. In general, it is these which precipitate the most antibody from the various sera. Qualitatively this applies as well to the three dextran samples, of which the least reactive is the partially hydrolyzed clinical sample, N150N. The polyglucose sample, P218-56,[‡] is comparable with this dextran in

average molecular size, but shows considerably lower serological activity except in Type XII antiserum.

According to the quantitative theory of the precipitin reaction which describes the mutual precipitation of antigen and antibody in terms of the multivalence of each reactant[5] one would expect a cross reaction to be more complete with increasing size of the antigen molecules. Larger numbers of reactive groupings should overcome hindrances to precipitation due to the less perfect fit of antigen and antibody than in the homologous reaction. However, Kabat and Berg[8] have shown that partially hydrolyzed clinical dextrans do not necessarily precipitate less antidextran from the sera of human beings than do undegraded samples. Further chemical data on both the polyglucose fractions and the various dextrans are therefore needed before the serological behavior of these substances can be fully explained.

Although the viscosities and reducing values of fractions D and E are almost identical their serological reactivities are not the same. Whether or not there are differences in the average sizes of the molecules in the two fractions, the presence of separable fractions with structural differences appears indicated.

The chemical basis for the quite extensive cross reactivity of some of the synthetic polyglucoses with antibodies specific for pneumococcal "C" polysaccharide and the type-specific polysaccharides of pneumococcal Types IX, XII, XX, and XXII is as yet uncertain. Since even the component sugars of these five bacterial carbohydrates are unknown it would seem premature to theorize until current attempts to supply at least a portion of the necessary knowledge are further advanced. Stacey has, however, pointed out that the specific polysaccharide of Type II pneumococcus and one of the dextrans contain branching points of glucose units with linkages in the 1, 4, and 6 positions.[14] If the reactions of the polyglucoses described above were due to the multiple recurrence of the same linkages one would expect the glycogens and amylopectins to give precipitates in antisera to the same types. The positive outcome of experiments to test this surmise will be described in a separate communication.

Summary.—Cross reactions of synthetic polyglucoses with antibodies to pneumococcal C-substance and type-specific polysaccharides are recorded and compared with similar reactions of several dextrans. A partial interpretation of the results is given.

* Carried out with the aid of the Harkness Research Fund and a grant from the Rockefeller Foundation. Generous gifts of antisera by the Bureau of Laboratories, New York City Department of Health, and the Division of Laboratories and Research of the Department of Health, State of New York, are also gratefully acknowledged.

† The writers are greatly indebted to Merck & Co., Inc., for these and other samples and data, and especially to Dr. Arthur J. Zambito of Merck & Co., Inc. for his many courtesies.

‡ Acknowledgment is due E. I. du Pont de Nemours & Co., Inc., for this sample and relevant data.

[1] Heidelberger, M., and Avery, O. T., *J. Exptl. Med.*, **38,** 73 (1923); **40,** 301 (1924).

[2] Hotchkiss, R. D., and Goebel, W. F., *J. Biol. Chem.*, **121,** 195 (1937); Reeves, R. E., and Goebel, W. F., *Ibid.*, **139,** 511 (1941).

[3] Haworth, N., Kent, P. W., and Stacey, M., *J. Chem. Soc.*, **1948,** 1220; the polysaccharide described *Ibid.*, 1211, cannot contain amino sugar, since it has been obtained nitrogen-free. Heidelberger, M., and Menzel, A. E. O., *J. Biol. Chem.*, **118,** 79 (1937).

[4] Reviewed in Heidelberger, M., *Bact. Rev.*, **3,** 49 (1939); *Chem. Rev.*, **24,** 323 (1939).

[5] Heidelberger, M., and Kendall, F. E., *J. Exptl. Med.*, **61,** 563 (1935).

[6] Heidelberger, M., Kabat, E. A., and Shrivastava, D. L., *Ibid.*, **65,** 487 (1937); Heidelberger, M., Kabat, E. A., and Mayer, M., *Ibid.*, **75,** 35 (1942).

[7] Sugg, J. Y., and Hehre, E. J., *J. Immunol.*, **43,** 119 (1942); cf. also Zozaya, J., *J. Exptl. Med.*, **55,** 353 (1932).

[8] Kabat, E. A., and Berg, D., *J. Immunol.*, May, 1953. In this paper it is shown, among other things, that two synthetic polyglucoses partially precipitate antibodies formed in man after injection of dextran. In part, also, in *Ann. N. Y. Acad. Sci.*, **55,** 471 (1952).

[9] Heidelberger, M., MacLeod, C. M., Markowitz, H., and Roe, A. S., *J. Exptl. Med.*, **91,** 341 (1950).

[10] Heidelberger, M., and Kendall, F. E., *Ibid.*, **53,** 625 (1931).

[11] Heidelberger, M., and Kendall, F. E., *Ibid.*, **61,** 559 (1935).

[12] Markham, R., *Biochem. J.*, **36,** 790 (1943); cf. also Kabat, E. A., and Mayer, M. M., *Experimental Immunochemistry*, C. C Thomas, Springfield, Ill., 1948.

[13] Pacsu, E., personal communication.

[14] Stacey, M., *Biochem. J.*, **53:** *Proc. Biochem. Soc.*, xiii (1953); *Endeavour*, **12,** 38 (1953).

---

*A NEW PARTICLE TYPE IN CERTAIN CONNECTIVE TISSUE EXTRACTS**

BY FRANCIS O. SCHMITT, JEROME GROSS, AND JOHN H. HIGHBERGER

THE DEPARTMENT OF BIOLOGY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE; THE DEPARTMENT OF MEDICINE, MASSACHUSETTS GENERAL HOSPITAL† AND HARVARD MEDICAL SCHOOL, BOSTON; AND THE UNITED SHOE MACHINERY CORPORATION, BEVERLY, MASSACHUSETTS

Communicated April 30, 1953

Chemical and electron microscopic studies of the in vitro reconstitution of fibrils from acid solutions of collagen have revealed that the form of the axial repeating structure in the fibril is markedly influenced by the physical chemical environment. This paper describes a new structure precipitated from collagen solutions, the organization of which is regarded as having fundamental significance for collagen structure.

Early studies[1–5] patterned after the classic Nageotte experiment[6] demonstrated that the addition of NaCl to dilute acid solutions of certain forms

of collagen, such as rat tail tendon or ichthyocol,[7] precipitate fibrils which, when viewed in the electron microscope, exhibit a very regular axial repeating period. Neutralization or addition of 0.1–0.2 *M* solutions of salts with monovalent cations precipitated fibrils with the axial period and detailed intraperiod fine structure characteristic of native collagen fibrils. At somewhat higher ionic strength the precipitated striated fibrils had periods of about one-third the normal, or 220 A. At still higher salt concentration the fibrils formed were structureless.[5, 8]

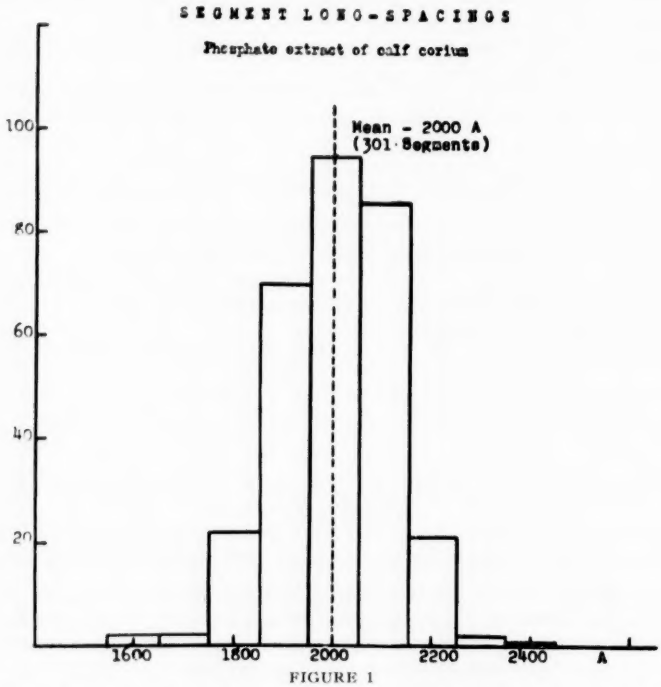


FIGURE 1

Frequency distribution curve of lengths of segments produced from phosphate extracts of calf corium.

Following the procedure for preparing "procollagen" described by Tustanovsky[9] and Orekhovich, *et al.*,[10] whereby skin and other tissues are extracted with acid citrate buffer and the extract dialyzed against water, the present authors found that the precipitate which forms on dialysis contains, besides typical collagen fibrils, a new fibril type.[11] This fibril, having an axial period ranging from 1800 to 3000 A, was called the "long-spacing" or "LS" form. To distinguish between this structure and the new struc-

ture to be described below, we shall refer to it as "fibrous long-spacing" or "FLS."

FLS could be produced also from acetic acid filtrates of fish swim bladder collagen (ichthyocol) by the addition of plasma α-1 acid glycoprotein followed by dialysis.[8] The acid filtrates contain about 0.1% collagen. At glycoprotein concentrations of about 0.02% or higher, FLS are formed almost exclusively. At lower concentrations (*ca.* 0.001%) of glycoprotein, chiefly collagen-type fibrils are reconstituted. At intermediate concen-

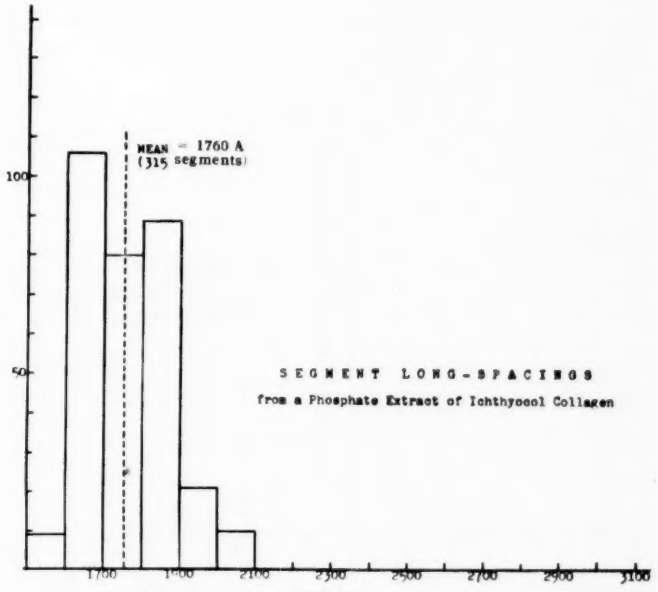


FIGURE 2

Frequency distribution curve of lengths of segments obtained from phosphate extracts of carp swim bladder tunic.

trations both types are formed. FLS can be converted into the collagen type by dissolving in 0.05% acetic acid and dialyzing against 1.0% NaCl.

Although glycoprotein is highly efficient in producing FLS and collagen, it is not unique in this property; a number of other substances will accomplish the same result. It was therefore concluded[12] that the components necessary for the formation of cross-striated fibrils, whether of the collagen or FLS type, are contained in the collagen extracted from the connective tissue. The precipitating substances seem to be relatively non-specific under the conditions of the experiment.[13]

Detailed intraperiod structure may be observed in LS fibrils. At least 14 bands of characteristic position and density have been observed in preparations stained with phosphotungstic acid (PTA). The pattern of banding (Fig. 5) is strikingly different from that of collagen (Fig. 6) not only in the relative positions of the bands but also in being non-polarized or symmetrical in the axial direction.

It is generally assumed that collagen is composed of protofibrils or polypeptide chains with sequences of amino acids running in the same direction in adjacent chains and with discontinuities in register, thus forming a cross-

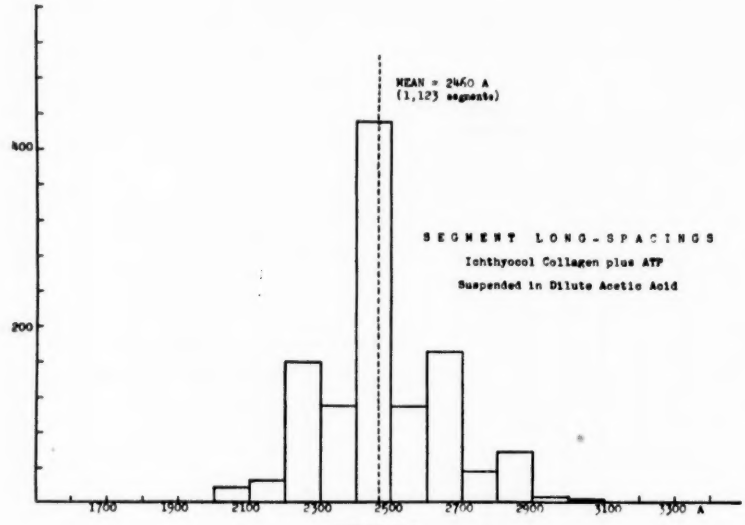


FIGURE 3

Frequency distribution curve of lengths of segments produced by the interaction of ATP and ichthyocol filtrate.

striated fibril with polarized structure. It is possible that the symmetry of FLS structure is produced by a packing of polarized chains oriented in both directions (antiparallel array).

Using 70,000 as the molecular weight of procollagen and 600 the number of amino acid residues, Bresler, *et al.*,[14] concluded that procollagen must be a highly coiled chain since they deduced that the length of the molecule is only 380 A. If the helix were uncoiled the molecule would have a length of the order of 2000 A. The suggestion has been made[15] that in LS fibrils the chains are uncoiled, being essentially in a β state. For this there is as yet no direct evidence. Although high-angle x-ray patterns of FLS prep-

arations displayed the features characteristic of normal collagen, more conclusive evidence is required to prove or refute this point of view.

Substantially all of the collagen in acid solutions of ichthyocol may be converted to FLS if the concentration of acid glycoprotein, added before dialysis, is sufficiently high. This suggests the possibility that thin particles having the length of the FLS period, i.e., about 2000 A, may occur in

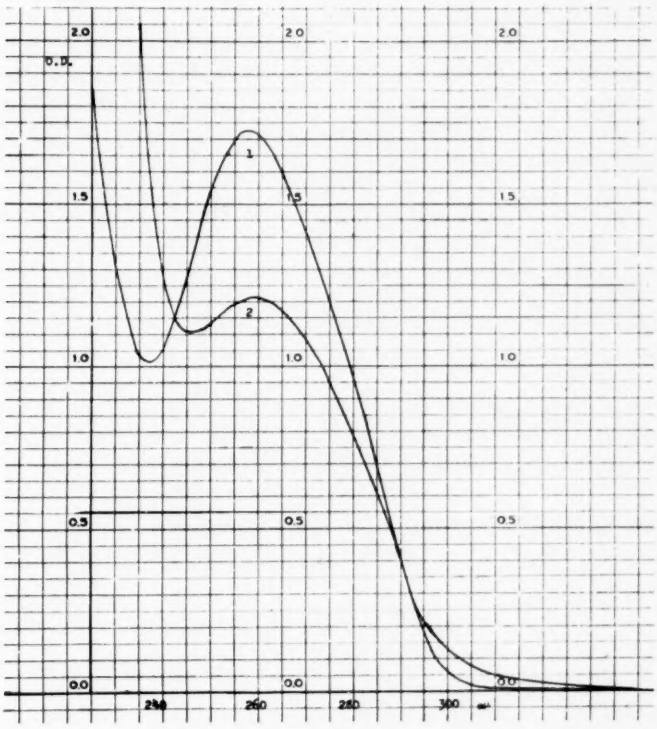


FIGURE 4

Ultraviolet absorption spectra of: 1. Phosphate extract of calf corium; 2. Phosphate extract of carp swim bladder tunic.

the collagen solutions or may be dissociated from a higher polymer. The fact that segments of the order of 2000 A in length were occasionally observed in "procollagen" solutions (Fig. 7) lends support to this view. Accordingly, efforts were made to isolate the material of which such elongate particles are composed.

Earlier experiments[8] indicated that FLS could be produced from mildly

alkaline phosphate extracts of skin and other connective tissues. Dialysis of such extracts produces an amorphous precipitate; extraction of this precipitate with citrate buffer (pH = 3.8, μ = 0.2), followed by dialysis, leads to the formation of FLS together with collagen-type fibrils.

A slight variation of this procedure led to the discovery of an entirely different and new long-spacing type. This involved the dialysis of a phosphate extract (pH = 8, μ = 0.4) against citrate buffer (pH = 4, μ = 0.2)—rather than the citrate extraction of the dialyzed precipitate of a phosphate extract. Shortly after the beginning of dialysis a white copious precipitate begins to form. In the electron microscope this material was seen to contain (besides variable amounts of unstructured material and thin filaments) a new component in the form of segments having a characteristic pattern of internal band structure (Fig. 8).

In figures 1 and 2 are shown the distribution curves of lengths of segments obtained from phosphate extracts of calf corium and carp swim bladder tunic. The curves differ with regard to the mean length and dispersion. The reasons for the differences are not yet clear; however, they serve at least to establish an order of magnitude for the segment length. The segment lengths clearly fall within the range of periodicity of the fibrous long spacings. These new structures will be called "segment long spacings" or "SLS."

Further characterization of the material precipitated as SLS required considerable purification. It proved very difficult to separate SLS from contaminants in skin extracts, either before or after precipitation. Far more satisfactory as starting material is the tunic of the carp swim bladder after preliminary extraction with dilute acetic acid or acid citrate buffer. In such preparations the finely filamentous contaminant is present only in

FIGURES 5–10 (P. 465)

Fig. 5. Long spacing fibril (FLS) produced by dialysis of a mixture of α-1 acid glycoprotein and ichthyocol filtrate. Note the detailed symmetrical intraperiod fine structure. PTA stained. Mag. 44,500×.

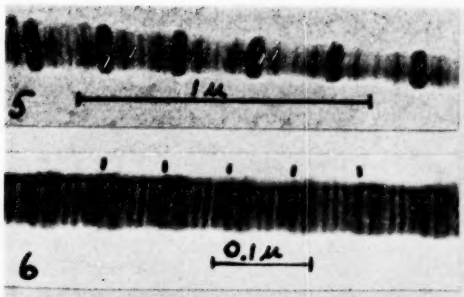
Fig. 6. Collagen fibril of cow hide showing characteristic polarized intraperiod banding. Stained with PTA. Mag. 150,000×.

Fig. 7. Segments of LS produced early in the dialysis of a citrate extract of newborn rat skin against water. Note similarity of intraperiod fine structure with that of FLS (Fig. 1). Mag. 24,000×.

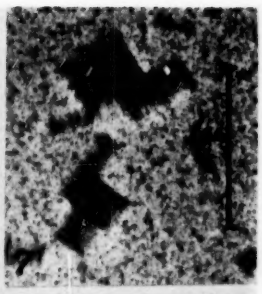
Figs. 8 and 8A. Segment long spacings obtained by dialysis of a phosphate extract of calf corium against citrate. Mag. 23,000×. Note detailed asymmetric fine structure particularly in the enlarged pair of segments joined end to end in 8A. Mag. 44,000×.

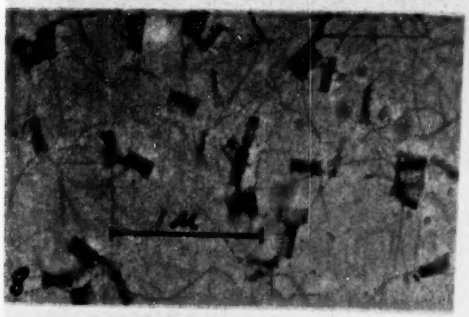
Figs. 9 and 9A. Segments obtained from phosphate extracts of carp swim bladder tunic which had been previously extracted with dilute acetic acid. PTA stained and lightly shadowed with chromium. Mag. 18,000×. More detailed fine structure is illustrated in 9A. (PTA alone). Mag. 80,000×.

Fig. 10. Segments, produced by the addition of ATP to an ichthyocol filtrate. PTA stained and lightly shadowed with chromium. Mag. 24,000×. 10A is a higher magnification. Mag. 49,000×.

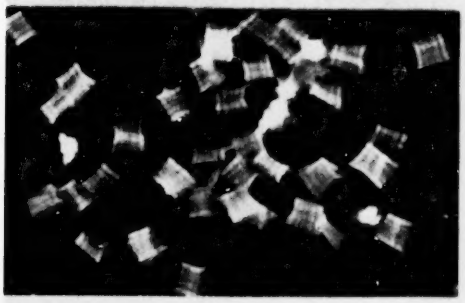

5
1 μ

6
0.1 μ

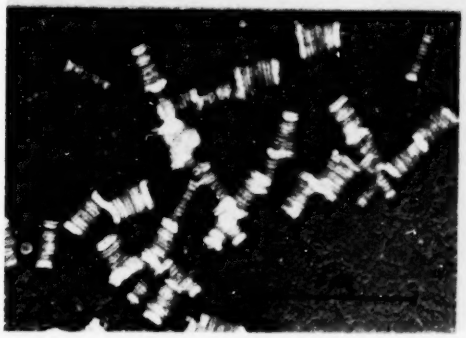

7


1 μ
8

0.5 μ
8A



0.2 μ
9A



very small amounts. An electron micrograph of purified SLS produced from a phosphate extract is shown in figure 9.

It will be noted that the SLS structure is polarized. This is most easily noticed by observing the interband region of very low density which occurs slightly off-center (Figs. 8A and 9A). As many as 18 bands have been observed in individual segments.

Phosphate extracts both of skin and of acid extracted swim bladder tunic manifest strong absorption at about 2600 A and a minimum at about 2375 and 2450 A (Fig. 4) suggesting the possibility that nucleic acid or nucleotides may play a role in the formation of SLS. It is true that certain solutions such as phosphate extracts of salt-reconstituted ichthyocol, which yield SLS upon dialysis against citrate buffer, show no absorption maximum at 2600 A. Nevertheless it seemed worth while to examine the efficacy of nucleic acid and various nucleotides in forming SLS from collagen solutions.

Yeast ribonucleic acid and DNA dissolved in 0.05 *M* citrate, pH = 4, and also in water, were added to ichthyocol filtrate to make concentrations ranging from 0.005 to 0.6%. Immediate precipitation of structureless filaments, but no SLS, occurred in all cases.

Adenosine-3-phosphoric acid, adenosine-5-phosphoric acid, uridylic, and guanylic acids in 0.2% concentration produced no precipitate. Hexametaphosphoric acid, sodium pyrophosphate, and the sodium salts of ADP and adenosine triphosphate produced fluocculent precipitates containing collagen-type or structureless fibrils immediately after mixing. No SLS or FLS were formed.

However, when adenosine triphosphoric acid (ATP), dissolved in 0.05% acetic acid, was added in 0.1 to 0.25% final concentration to ichthyocol filtrate, precipitation started almost immediately and continued for several hours in the cold. These precipitates are composed exclusively of SLS although some very finely filamentous material is present. Their structure (Figs. 10 and 10A) is essentially identical with that of SLS obtained from phosphate extracts of skin and swim bladder. Figure 3 is a distribution curve of segment length in such a preparation of ATP-induced SLS. There is considerable variability in segment width from one preparation to the next and ofttimes within the same preparation as shown in figures 8

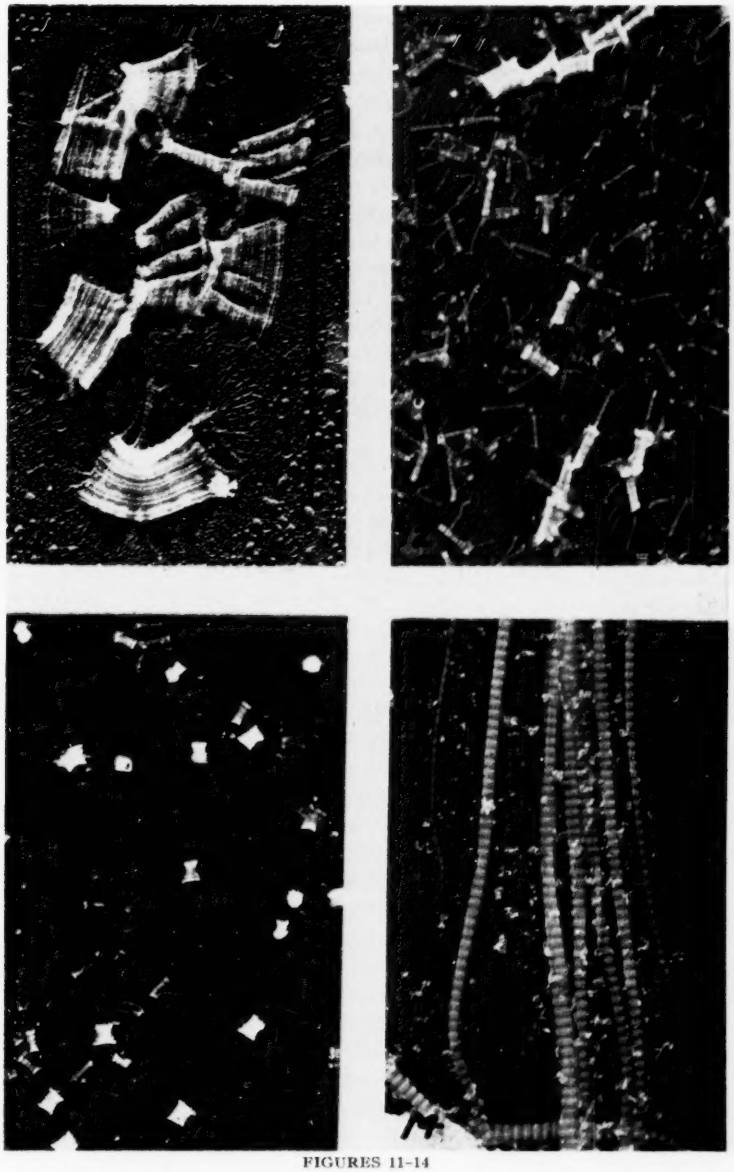
FIGURES 11–14

Fig. 11. ATP-produced segments washed in dilute acetic acid. Fraying indicates the parallel array of filamentous components composing the segment. Mag. 42,000×.

Fig. 12. ATP-produced segments revealing great variability in width and also emphasizing the fibrous nature of the components. Mag. 21,000×.

Fig. 13. ATP-induced SLS showing several standing on end. Calculation from shadow length indicates a height nearly equal to the length of those segments lying flat. Their three dimensional quality is also apparent. Mag. 12,000.

Fig. 14. SLS transformed to collagen after dissolving in NaCl and dialyzing against the same salt. A small amount of amorphous debris is usually present. Mag. 27,000×.



FIGURES 11-14

and 12. Frequently one observes segments standing on end, as in figure 13, and it is obvious that these structures are not flat ribbons but have significant thickness. This is also evident from the shadows cast by those segments lying flat. Preliminary study indicates that SLS is very highly hydrated. Application of the freeze drying technique of Williams[16] may provide interesting information on the shape of these structures.

Washing the precipitate several times with fresh 0.1% ATP in 0.05% acetic acid further removed the earlier described contaminant and provided a highly purified preparation of SLS. When such purified SLS is exposed to acetic acid in the same concentration in which they were formed (0.05%, pH = 3.5) the segments begin to disintegrate. This process is in itself of considerable interest for, when followed with the electron microscope, it lends strong support to the view that the segments are composed of a parallel packing of very thin fibrous particles of uniform length (Fig. 11). It may also be seen that the banded structure is still evident in some of the thin segment frays.

For present purposes it seems reasonable to assume that the chains in the SLS are in parallel array and that ATP molecules tend to bond neighboring chains laterally. More evidence is needed before a definitive suggestion can be made regarding the specific nature of this bonding. Preliminary estimates indicate that SLS preparations which had been washed with acid may contain as much as 12% of ATP which would correspond to about one mole of ATP per 3278 grams of protein.

Particularly pertinent is the question of the relationship between collagen, fibrous LS, and segment LS. Can they be converted one into the other and, if so, by what means? Experiments outlined below show that all of these forms are interconvertible.

Conversion of SLS to Collagen.—When a pure preparation of SLS is dissolved in 0.17 *M* NaCl and dialyzed at room temperature against NaCl, a heavy fibrous precipitate forms which is composed entirely of well-structured collagen fibrils with the characteristic collagen banding and a small amount of granular debris (Fig. 14). The conversion to collagen is accelerated at temperatures up to 37°C.

Conversion of Collagen to SLS.—Collagen-type fibrils with characteristic banding may be obtained by dialyzing an acid filtrate of ichthyocol against a 1% NaCl solution. This precipitate of collagen fibrils, after removal by centrifugation, was dialyzed exhaustively in the cold against 0.05% acetic acid. The precipitate dissolved to form a tenuous gel. This was diluted to the original volume with 0.05% acetic acid and ATP added to make a concentration of 0.2%. The precipitate which formed was nearly pure SLS.

Conversion of Fibrous LS to Segment LS.—LS fibrils, produced by dialysis of a mixture of ichthyocol filtrate and acid glycoprotein (procollagen), may

be dissolved in acetic acid. Dialysis of such a solution against 1% NaCl produced collagen fibrils while addition of ATP produced SLS.

These results demonstrate that the various forms of fibrous "collagens" may be freely converted, one into the other. The likelihood that thin fibrous protein particles about 2200 A long, or polymers thereof, exist in various extracts of connective tissue is indicated both by the action of glycoprotein, to form the symmetrical fibrous LS, and by the action of ATP, to form the polarized, asymmetric segment type of LS.

Preliminary chromatographic analyses reveal that the amino acid pattern of SLS protein resembles that of collagen fairly closely. Tyrosine values are low and this amino acid may actually be absent. Hydroxyproline is present in ichthyocol SLS in a concentration approximately 90% of that found in ichthyocol collagen. After allowance is made for the ATP nitrogen, the nitrogen content of the SLS protein was estimated to be 18.2%, in good agreement with that of collagen. Films of SLS give a high-angle x-ray pattern characteristic of collagen.

Neither fibrous nor segmental LS have thus far been observed in tissues. This may be due to the relatively high ionic strength of tissue fluids and to the fact that the long particles may not exist in the uncombined state to any appreciable degree either in the ground substance or in the formed fibrils of the connective tissue. This problem is being investigated.

The possible range of conditions under which SLS may form has not yet been determined. It would be premature to implicate ATP itself or other nucleotides specifically in their formation from native tissue extracts. The presence of purine- or pyrimidine-containing compounds in the phosphate extracts may conceivably be coincidental. It is quite possible that the long protein particles may be capable themselves of aggregating in the manner characteristic of SLS when the physical chemical environment is appropriate. The possible relationship of these particles to collagen precursors is at present a matter for speculation.

Elongate particles or molecules having lengths of the order of 1500 to 3000 A have been deduced to exist in tropomyosin,[17] L-myosin,[18] paramyosin,[19] and keratin.[20] It is possible that application of the type of technique described above may permit direct electron microscopic visualization of the elongate particles through their reaction with a compound like ATP which may unite chains of similar length and constitution laterally and form either segmental or fibrous LS structures, as in the case of collagen.

Acknowledgment.—The authors wish to thank Mrs. Zelda Sokal, Mr. R. A. Whitmore, Mr. L. S. Allen, and Mr. J. W. Jacques for their technical assistance.

* This investigation was supported in part by grants from the Eli Lilly Company to

the Massachusetts Institute of Technology and from the U. S. Public Health Service to the Massachusetts General Hospital.

† This is publication No. 141 of the Robert W. Lovett Memorial Foundation for the Study of Crippling Disease, Department of Medicine, Massachusetts General Hospital and Harvard Medical School.

[1] Schmitt, F. O., Hall, C. E., and Jakus, M. A., *J. Cell Comp. Physiol.*, **20,** 11–33 (1942).

[2] Bahr, G., *Exptl. Cell Research*, **1,** 603–606 (1950).

[3] Vanamee, P., and Porter, K. R., *J. Exptl. Med.*, **94,** 255–268 (1951).

[4] Noda, H., and Wyckoff, R. W. G., *Biochem. et Biophys. Acta.*, **7,** 494–506 (1951).

[5] Gross, J., *Proc. Josiah Macy Jr. Foundation;* Conference on Metabolic Interrelationships, **4,** 32–57 (1952).

[6] Nageotte, J., and Guyon, L., *Arch. Biol.*, **41,** 1–35 (1930).

[7] Fauré-Fremiet and Garrault, H., *Arch. Anat. mic.*, **33,** 81–94 (1937).

[8] Highberger, J. H., Gross, J., and Schmitt, F. O., PROC. NATL. ACAD. SCI., **37,** 286–291 (1951).

[9] Tustanovsky, A. A., *Biokhimiya*, **12,** 285–290 (1947).

[10] Orekhovich, V. N., Tustanovsky, A. A., Orekhovich, K. D., and Plotnikova, N. E., *Biochemiya*, **13,** 55–60 (1948).

[11] Highberger, J. H., Gross, J., and Schmitt, F. O., *J. Am. Chem. Soc.*, **72,** 3321 (1950).

[12] Gross, J., Highberger, J. H., and Schmitt, F. O., *Proc. Soc. Exptl. Biol. Med.*, **80,** 462–465 (1952).

[13] It should be noted, however, that a wide variety of proteins and other high polymers will not reproduce this phenomenon.

[14] Bresler, C. E., Finogenov, P. A., and Frenkel, S. Y., *Repts. Acad. Sci. U.S.S.R., Moscow*, **72,** 555–558 (1950).

[15] Bear, R. S., *Adv. Protein Chem.*, **7,** 69–160 (1952).

[16] Williams, R. C., *Biochem. et Biophys. Acta*, **9,** 237–239 (1952).

[17] Bailey, K., "The Muscle. A Study in Biology and Pathology," *L'Expansion Sci. Francaise*, 31–46 (1950).

[18] Weber, H. H., *Adv. Protein Chem.*, **7,** 161–252 (1952).

[19] Hodge, A. J., PROC. NATL. ACAD. SCI., **38,** 850–855 (1952).

[20] Mercer, E. H., and Olofsson, B., *J. Polymer Sci.*, **6,** 671–680 (1951).

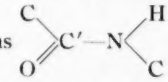
HYDROGEN BONDED HELICAL CONFIGURATIONS OF THE POLYPEPTIDE CHAIN

BY JERRY DONOHUE

THE MEDICAL RESEARCH COUNCIL UNIT FOR THE STUDY OF THE MOLECULAR STRUCTURE OF BIOLOGICAL SYSTEMS, CAVENDISH LABORATORY, CAMBRIDGE, ENGLAND

Communicated by Linus Pauling, April 13, 1953

Pauling and Corey[1–3] have formulated two helical configurations for polypeptide chains in which: (1) the residues have the dimensions deduced by consideration of the results of crystal structure studies in amino acids and related compounds,[4] (2) all residues are equivalent, and (3) each residue forms an N—H···O hydrogen bond of length about 2.75 Å in

which the nitrogen-oxygen vector is not more than 30° from the nitrogen-hydrogen vector. An important feature of the residues is the planarity of the atoms  , in each. The two helices were termed by Pauling and Corey the α-helix and the γ-helix. A convenient system for assigning an unambiguous designation to helical polypeptide structures with equivalent residues is that of Bragg, Kendrew and Perutz.[5] In this system each helix is designated by two numbers, S_R, where S denotes the screw symmetry and R the number of atoms in the hydrogen bonded

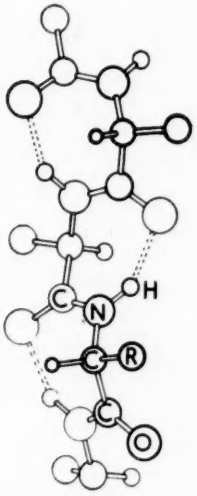


FIGURE 1

The 2.2_7 helix.

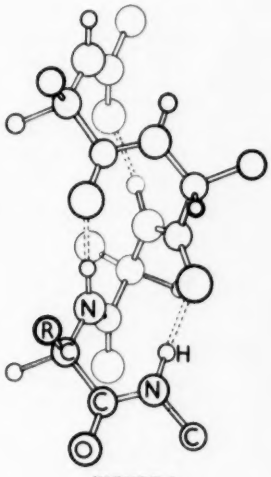
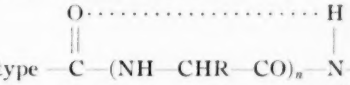


FIGURE 2

The 3.0_{10} helix.

ring. The numbers R are of two families, being equal to $3n + 4$ for rings of the type  , and $3n + 5$ for rings of the type . The α-helix is then 3.6_{13}, and the γ-helix is 5.1_{17}. For a helix of given R, the conditions postulated by Pauling and Corey in some cases render the construction of a stereochemically satisfactory helix impossible, otherwise they determine S within rather narrow limits.

Pauling and Corey stated that the α and γ helices were the only solutions to the problem which conformed with their basic postulates. Upon systematic examination of all structures with $R < 17$ it has been found, however, that there are four additional helical configurations of polypeptide chains which also conform to these postulates, to the degree described below. One of these, having designation 2.2_7 and an axial translation of 2.75 Å per residue, resembles the "α_{II}" structure first suggested by Huggins.[6] This is a structure which has been the subject of discussion in a number of papers since then, not all of which need be referred to here. The relation of the present version with some of those put forward previously is discussed below.

TABLE 1

ATOMIC COORDINATES FOR THE 2.2_7, 3.0_{10}, AND 4.3_{14} HELICES

	x	y	z	ρ	φ
		2.2_7			
C	0.16 Å	1.31 Å	0.00 Å	1.32 Å	0.0°
N	−0.15	0.52	1.20	0.54	−23.1
H	−0.89	0.80	1.80	1.20	−55.0
C′	0.54	−0.54	1.48	0.76	128.0
O	1.46	−0.92	0.77	1.73	115.2
C*	0.16	−1.31	2.75	1.32	166.0
βC-1	1.09	−2.53	2.89	2.75	149.7
βC-2	−1.27	−1.84	2.61	2.24	209.6
		3.0_{10}			
C	0.93 Å	1.61 Å	0.90 Å	1.86 Å	0.0°
N	0.86	0.62	1.09	1.06	24.2
H	0.69	0.92	2.03	1.15	6.9
C′	1.02	−0.65	0.82	1.21	92.5
O	1.23	−1.07	−0.33	1.63	101.0
C*	0.93	−1.61	2.00	1.86	120.0
βC-1	2.31	−2.01	2.55	3.06	101.0
βC-2	0.23	−2.90	1.52	2.91	145.5
		4.3_{14}			
C	2.01 Å	1.78 Å	0.00 Å	2.68 Å	0.0°
N	2.05	−0.33	1.14	2.08	50.7
H	2.07	0.20	1.97	2.08	36.0
C′	2.05	0.27	−0.02	2.07	34.0
O	2.07	−0.33	−1.08	2.10	50.6
C*	2.01	−1.78	1.20	2.68	83.0
βC-1	2.66	−2.38	−0.05	3.57	83.3
βC-2	2.66	−2.38	2.45	3.57	83.3

The existence of a second, π-helix, has been pointed out by Low and Baybutt;[7] this helix has the designation[8] 4.4_{16} and an axial translation of 1.15 Å per residue. The two other structures have designations 3.0_{10}

and 4.3_{14}, and axial translations of 2.00 Å and 1.20 Å per residue respectively. Atomic coordinates for the 2.2_7, 3.0_{10}, and 4.3_{14} helices are presented in table 1. Drawings for the four structures are shown in figures 1, 2, 3, and 4. (The coordinates used in preparing the drawing for the 4.4_{16} helix were determined independently in this work.)

Studies were not made of helices with R greater than 17, because an unfavorable situation with regard to van der Waals attraction is expected for them. The two lower members of the $R = 3n + 5$ family may be rejected because of the great strain energies associated with even their most favorable formulation. For $R = 8$ and planar residues, the shortest possible N···O distance is 4.2 Å. This can be reduced to an acceptable

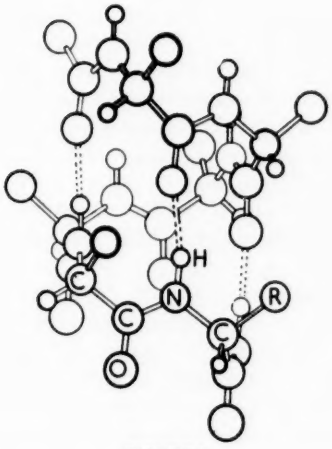


FIGURE 3

The 4.4_{16} helix.

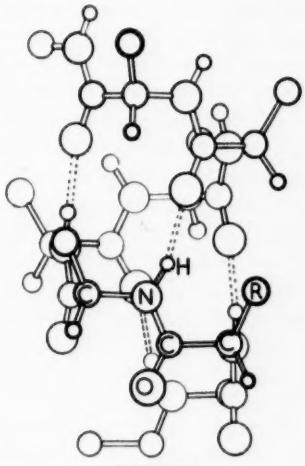


FIGURE 4

The 4.3_{14} helix.

value by rotation of each residue about the C′—N bond by about 45°, thus introducing a strain energy of 15 kcal. per mole per residue. This amount could be reduced somewhat by decreasing some of the bond angles, but only at the expense of additional strain due to bond deformation. The structure with $R = 11$ can be constructed only by allowing, first, a rotation through an angle δ of the peptide groups about the C′—N bonds, and second, a decrease of the angle C′—C—N from 110°. While preserving a satisfactory hydrogen bond system it is found that minimum strain energy occurs at about $\delta = 20°$, angle C′—C—N = 97°. These values correspond to total strain energy of more than 5 kcal. per mole per residue, so great as to render this helix unacceptable also. A 3.6_{11} helix has been

described recently by Huggins,[10] and atomic coordinates for it have been published.[11] This helix involves great deviation from the structural postulates; Pauling and Corey[12] have discussed the amount of the strain energy.

Of the numerous papers in which the α_{II} ribbon (2.2_7 helix) has been discussed, only one, that of Bragg, Kendrew and Perutz,[5] gives coordinates for the atoms. It is therefore impossible to test the other models for their conformity with known structural principles, since sketches or photographs of models are not sufficient. Moreover, it is noteworthy that all former models possess an exact twofold screw axis, whereas with this restriction it is not possible to construct a satisfactory structure. For example, if the coordinates for the hydrogen atom are obtained by measurement of the drawings of the 2_7 structure of Bragg, Kendrew and Perutz,[5] the N—H···O angle differs from a straight angle by 66°, and that the angle between the N—H and N···O vectors is 48°. A second example of an unacceptable structure for the α_{II} ribbon (with a twofold screw axis) is that of Robinson and Ambrose,[9] who investigated various chains with their space filling models. Although no coordinates were given, it is possible to derive from the data they do quote that the N—H···O angle differs from a straight angle by about 60°, the angle between the N—H and N···O vectors is about 40°, and, moreover, the N···O distance is about 2.56 Å, more than 0.1 Å shorter than any reliable value so far reported.

A 3_{10} helix which deviates widely from the structural postulates has been put forward previously by Huggins.[6] Here again it is difficult to determine the precise characteristics of this model, since no atomic coordinates were given, but it is clear from the drawings of it that the peptide groups deviate so far from planarity that it must be considered unacceptable. Furthermore, 5.1 Å was given as the axial repeat for 3 residues, as compared with 6.0 Å for the present structure, and it has been found that the axial repeat of the 3.0_{10} helix may not be shortened by as much as 0.2 Å (for 3 residues) without either leading to an improbably short N—H···O distance or introducing other undesirable features.

A helix with $R = 14$ does not appear to have been described before. Bragg, *et al.*,[5] state that for 3_{14} and 4_{14} there are "no possible structures," but these workers restricted themselves to integral values of S, the screw symmetry.

There are several short distances involving the two possible positions for the β-carbon atoms. In the 2.2_7 helix the distance from β-carbon-2 to the next oxygen atom above it is only 2.6 Å, and in the 4.3_{14} and 5.1_{17} (γ) structures the distance from β-carbon-2 to the next oxygen atom below is even shorter, a situation which may eliminate these choices for the position of the β-carbon atoms. In the 3.6_{13} (α), and 4.4_{16} struc-

tures the distances from β-carbon-2, and in the 2.2_7 structure the distance from β-carbon-1, to the next oxygen below is 2.7 Å, not far from the C···O distances observed in acetylglycine,[13] glutamine,[14] glycylglycine,[15] and glycylasparagine,[16] where C···O distances of 2.69 Å, 2.76 Å, 2.78 Å, and 2.81 Å respectively, are found. All these are close to what is expected when the βC—C bond, or C—N bond in the case of the peptides, is *cis* to the C′—O bond. (In the 3.0_{10}, 4.3_{14}, and 5.1_{17} helices this distance is 2.8 Å or more, since the orientation of the βC—C bonds is 20° or greater from *cis*.) A preference between the two β-carbon positions cannot, therefore, be made on this basis. The position of the *R* groups in figures 1 to 4 corresponds to β-carbon-1.

It is a matter of some interest to compare the relative stabilities of the six helices. Five features of the structures will be examined: (1) distortion from planarity of the peptide residues; (2) deformation in bond angles; (3) van der Waals contacts; (4) the orientations about the single bonds, C—C′ and C—N; and (5) nature of the hydrogen bond. In all models the covalent bond distances are within ±0.02 Å of the predicted values,[4] which in any case, are accurate to only about that amount.

Quantitative estimates of the strain energies are made, where possible, as follows: (1) the strain energy due to a rotation of a peptide residue about its C′—N bond is given by the formula[17, 18] $A \sin^2 \delta$ where δ = the angle of rotation and A = 30 kcal. per mole; (2) for bending of single covalent bonds, the strain energy is calculated by assuming that the bond energy is proportional to the strength of the bond orbital in the bond direction;[12] thus, for the case of distortion of a bond angle this leads to the formula $1.5\ B \sin^2 \delta'/4$ where δ' is the deviation of the bond angle from the value expected for maximum bond strength and B is the sum of the bond energies of the two bonds forming the strained angle; (3) quantitative estimation of van der Waals forces, either repulsive or attractive, has not been attempted but rather the van der Waals contacts in the different structures are listed for comparison with each other and with other known structures; (4) according to Pauling and Corey[20] there are expected to be six favored orientations about each of the C—C′ and C—N bonds, each differing by 60°. The energy difference between a most favorable and most unfavorable position has been estimated at about one kcal. per mole, and may be calculated for intermediate positions by assuming that the potential function is a sine wave; (5) the energy[2] of an N—H···O hydrogen bond is of the order of 8 kcal. per mole if the angle between the N—H and N···O vectors, i.e. <H—N···O, is close to 0°. For values of this angle larger than 30° the energy is assumed to be negligibly small. It is not known in what way the energy decreases as the angle increases.

The results of the above considerations are presented in table 2. The hydrogen bonds are seen to be satisfactory in all six, the greatest deviation

TABLE 2

ESTIMATED RELATIVE STABILITIES OF SIX HELICAL POLYPEPTIDE STRUCTURES

	2.2_7	3.0_{10}	3.6_{13} (α)	4.3_{14}	4.4_{16} (π)	5.1_{17} (γ)
Deviation of residues from planarity	None	None	None	3° 0.1 kcal.[a]	None	None
Deformation of bond angles	None	None	None	C′—C—N = 100° 0.3 kcal.	C′—C—N = 114° 0.1 kcal.	None
Van der Waals contacts of 3.2 Å or less[b]	N···N = 3.0 Å C···O = 3.2 C···C = 3.1	N···N = 2.7 Å N···O = 2.7 C···O = 2.7 C···C = 2.9 C···N = 2.9	N···N = 2.9 Å N···O = 3.0 C···O = 2.7 C···C = 2.9	N···N = 3.0 Å N···O = 3.0	N···N = 3.0 Å C···O = 2.9 C···C = 3.0	N···N = 3.2 Å N···O = 3.2
Orientation about:						
C—N bond	15° from favorable 0.5 kcal.	10° from favorable 0.2 kcal	Favorable	25° from favorable 1 kcal.	10° from favorable 0.2 kcal.	30° from favorable 1 kcal.
C—C′ bond	Favorable	20° from favorable 0.8 kcal.	Favorable	25° from favorable 1 kcal.	10° from favorable 0.2 kcal.	30° from favorable 1 kcal.
Nature of hydrogen bond:						
Distance N···O	2.76 Å	2.81 Å	2.86 Å	2.80 Å	2.83 Å	2.77 Å
Angle H—N···O	24°	3°	4°	11°	11°	17°

[a] Energy differences are given in kcal. per mole per residue.

[b] The van der Waals distances may be compared with the contact C···O = 2.7 Å which occurs within a single residue in *all* the structures, with intramolecular contacts N···O of 2.64 Å to 2.69 Å in amino acid crystals,[19] and the intramolecular C···O distance of from 2.67 Å to 2.78 Å in the peptides cited previously.

from the ideal linear configuration being shown by the 2.2_7 structure. Although an angle of as much as 24° between the N—H and N···O vectors is not without precedent, it may be significant that this value is about as good as can be achieved for the 2.2_7 helix, whereas in crystal structures in which a similar deviation occurs there is always at least one other hydrogen bond of a more satisfactory nature. Omitting from consideration the van der Waals forces and hydrogen bond energies, it is seen that if the quantitative estimates are correct, the stability of the five structures is in the order 3.6_{13}, $4.4_{16}(0.5)$, $2.2_7(0.5)$, $3.0_{10}(1.0)$, $5.1_{17}(2.0)$, and $4.3_{14}(2.4)$, where the numbers in parentheses give the estimated instability in kcal. per mole per residue. Absence of appreciable van der Waals attractive forces in the 5.1_{17} helix would lead to additional instability.[18, 20] This effect also operates to a lesser extent with respect to the 2.2_7, 4.3_{14}, and 4.4_{16} structures, although this disadvantage might be overcome by the 2.2_7 helix because its flat character may enable two of them to

TABLE 3

STRUCTURAL CHARACTERISTICS OF HELICAL POLYPEPTIDE STRUCTURES

ATOMS IN RING	RESIDUES PER TURN	AXIAL TRANSLATION PER RESIDUE	ESTIMATED MINIMUM INSTABILITY PER MOLE PER RESIDUE
7 (α_{II} ribbon)	2.2	2.75 Å	0.5 kcal.
10	3.0	2.00	1.0
13 (α)	3.6	1.50	0.0
16 (π)	4.4	1.15	0.5
8	No possible structure		>15.0
11	No possible structure		> 5.0
14	4.3	1.20	2.4
17 (γ)	5.1	0.98	2.0

pack well side by side. The 3.0_{10} helix is the tightest of the six, and it is possible that the contacts are so close that the repulsion forces counteract the attractive forces to such an extent as to render it unstable, although none of the distances is unreasonably short in comparison with known structures. Van der Waals radii, particularly for intramolecular distances, are not accurate enough to settle the question. It seems likely that the strain associated with the 4.3_{14} helix is great enough to render its occurrence improbable. The values 3° for the twist in the peptide group and 100° for the angle C′—C—N represent the configuration of minimum strain energy for simultaneous variation of these parameters. It may be pointed out, on the other hand, that angles at a singly bonded carbon atom of as low as 104° have been observed in threonine[21] and as high as 115° in glutamine,[14] values not far from those in the 4.3_{14} and 4.4_{16} structures. It may well be possible that a combination of a favorable hydrogen bond situation plus unknown external influences may allow polypeptide chains to overcome small strains and to fold in one or another of the

configurations which, at first glance, seem to be rather less reasonable than the 3.6_{13} (α) helix.

It is interesting that the four most stable helices are all members of the family $R = 3n + 4$, and it is possible to pass from one structure to the next by moving the hydrogen bonds along the chain. The structural characteristics of the various chains of both families are summarized in table 3.

Radial distribution functions and x-ray scattering characteristics of these helical structures will be discussed in a subsequent paper.

Acknowledgment.—This work was supported by a Grant of a Fellowship from the John Simon Guggenheim Memorial Foundation.

[1] Pauling, L., and Corey, R. B., *J. Am. Chem. Soc.*, **72,** 5349 (1950).
[2] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, **37,** 205 (1951).
[3] Pauling, L., and Corey, R. B., *Ibid.*, **37,** 235 (1951).
[4] Corey, R. B., and Donohue, J., *J. Am. Chem. Soc.*, **72,** 2899 (1950).
[5] Bragg, W. L., Kendrew, J. C., and Perutz, M. F., *Proc. Roy. Soc.*, **A203,** 321 (1950).
[6] Huggins, M. L., *Chem. Rev.*, **32,** 195 (1943).
[7] Low, B. W., and Baybutt, R. B., *J. Am. Chem. Soc.*, **74,** 5806 (1952).
[8] Private communication from Dr. Low.
[9] Robinson, C., and Ambrose, E., *Trans. Far. Soc.*, **48,** 854 (1952).
[10] Huggins, M. L., *J. Am. Chem. Soc.*, **74,** 3963 (1952).
[11] Huggins, M. L., *Ibid.*, **74,** 3963 (1952).
[12] Pauling, L., and Corey, R. B., *Ibid.*, **74,** 3964 (1952).
[13] Carpenter, G. B., and Donohue, J., *Ibid.*, **72,** 2315 (1950).
[14] Cochran, W., and Penfold, B. R., *Acta Cryst.*, **5,** 644 (1952).
[15] Hughes, E. W., and Moore, W. J., *J. Am. Chem. Soc.*, **71,** 2618 (1949).
[16] Katz, L., Pasternak, R., and Corey, R. B., *Nature*, **170,** 1066 (1952).
[17] Pauling, L., and Corey, R. B., these PROCEEDINGS, **37,** 251 (1951).
[18] Corey, R. B., and Pauling, L., *Proc. Roy. Soc.*, **B141,** 10 (1953).
[19] Donohue, J., and Trueblood, K. N., *Acta Cryst.*, **5,** 419 (1952).
[20] Pauling, L., and Corey, R. B., these PROCEEDINGS, **37,** 729 (1951).
[21] Shoemaker, D. P., Donohue, J., Schomaker, V., and Corey, R. B., *J. Am. Chem. Soc.*, **72,** 2328 (1950).

THE EFFECT OF A SUPPRESSOR ON ALLELIC INOSITOLLESS MUTANTS IN NEUROSPORA CRASSA*

BY NORMAN H. GILES AND C. W. H. PARTRIDGE

OSBORN BOTANICAL LABORATORY, YALE UNIVERSITY

Communicated by P. R. Burkholder, March 30, 1953

Extensive analyses of spontaneous and radiation-induced reversions in a series of independently induced allelic inositolless mutants of *Neurospora crassa* have revealed that most such inositol-independent types are the result of reverse mutation.[1, 2] Furthermore, evidence was obtained that the inositolless mutants, which are phenotypically and biochemically indistinguishable,[1, 3] can be separated into several groups on the basis of quantitative differences in their spontaneous or induced reverse mutation rates. In addition to typical revertants to inositol-independence by reverse mutation, occasional phenotypically atypical revertants were recovered which, on crossing to wild type, segregated inositolless cultures in random ascospore isolation tests, indicating the probable occurrence of occasional suppressor mutations. The present paper deals with a more detailed genetic analysis of one such suppressor type. In addition, tests of the effect of this suppressor on other inositolless mutants have been made. These studies demonstrate that the suppressor is highly specific, since it apparently does not suppress any of the other mutants. Consequently, these results serve to establish a further difference between the suppressed inositolless mutant and all other inositolless mutants tested to date.

Origin, Genetic Analysis, and Biochemical Characteristics of the Suppressor Mutant.—The mutant described in these studies appeared following platings on minimal (inositol-free) medium of microconidia of inositolless strain 37401 (Stanford number) which had been exposed to ultraviolet. The mutant was initially noted as growing more slowly on minimal than the usual reverse mutation types. Even though the reversion had arisen in a microconidial strain, the mutant was next back-crossed to mutant 37401 and an inositol-independent isolate recovered in order to be certain that all subsequent tests would be performed with a homocaryotic strain. This extracted inositol-independent culture was then crossed to a wild-type inositol-independent strain and serial ascospore isolations made and tested, with the results indicated in table 1. The recovery of inositolless types as indicated by the presence of 4:4 and 6:2 ascus segregations demonstrates that the mutant being studied did not arise from reverse mutation, but rather was the result of mutation at some other locus. Such types in Neurospora have been designated, following the Drosophila terminology, suppressor mutants.[4]

The recovery of the three segregating types indicated in table 1 pro-

vides material for comparing the growth characteristics of the suppressor with the original inositolless mutant and with wild type (inositol-independent). These also furnish cultures in which the suppressor is present with the wild type allele at the inositolless locus. Growth tests of a series of such cultures have been carried out (at 25°C.) in 125-ml. Erlenmeyer flasks, using minimal medium and minimal supplemented with 4 μg. inositol per ml. Pads were removed at 3, 6, and 8 days for dry weight determinations, three flasks being used to provide an average value for each determination. Growth curves for asci exhibiting 6:2, 8:0, and 4:4 segregations are presented in figures 1, 2, and 3, respectively. The most instructive segregation is that shown in figure 1, the 6:2 type. In all instances the indicated genotypes were confirmed by appropriate crossing tests. It is immediately clear that all four genotypes are phenotypically distinguishable. It is also evident from the growth response in the culture from spore 16.3 that the suppressor only partially suppresses the effect of the inositolless mutation in 37401, since growth of the suppressor in the absence of inositol is considerably less than that obtained in the inositol-independent wild type. Further, added inositol markedly stimulates the

TABLE 1

RESULTS OF A CROSS OF REVERSION (HOMOCARYOTIC CULTURE EXTRACTED FROM A BACK CROSS TO 37401) WITH WILD TYPE (INOSITOL-INDEPENDENT).

ASCUS SEGREGATION PATTERN— INOSITOL-INDEPENDENT:INOSITOLLESS	NUMBER OF COMPLETE ASCI TESTED
4:4	2
6:2	8
8:0	2

growth of the suppressor. The growth response of the suppressor tends to be rather variable when different stocks are compared in simultaneous experiments (e.g., compare spore cultures 6.1, 6.3, and 16.3), perhaps due to segregating modifiers. However, there is also a rather considerable variability when the same stock is used in successive experiments. But in no instance has the growth of a suppressor strain closely approached that of wild type.

In addition to the suppressor which is segregating with inositolless mutant 37401 in spore 16.3, two additional cultures of the expected types, inositolless 37401 alone (spore 16.5) and inositol-independent (spore 16.7), are present. The fourth culture, from spore 16.1, in which the suppressor is present with wild type, is phenotypically distinct and exhibits a growth pattern which could not have been predicted in advance. It is evident, however, from culture 16.1 as well as from cultures 15.5 and 15.7 (figure 3), that the presence of the suppressor results in a marked delay in the growth on minimal of the wild type strain and that this delay is not overcome by the addition of inositol.

Microbiological assays have indicated that the suppressor culture is actually synthesizing *meso*-inositol, but apparently at a much reduced rate compared with the inositol-independent wild type. Quantitative assays have been performed with mutant 37401 of *Neurospora crassa* and with *Ashbya gossypi*, for the first of which a marked specificity of response to *meso*-inositol apparently exists.[5] Following 21 days' growth on minimal,

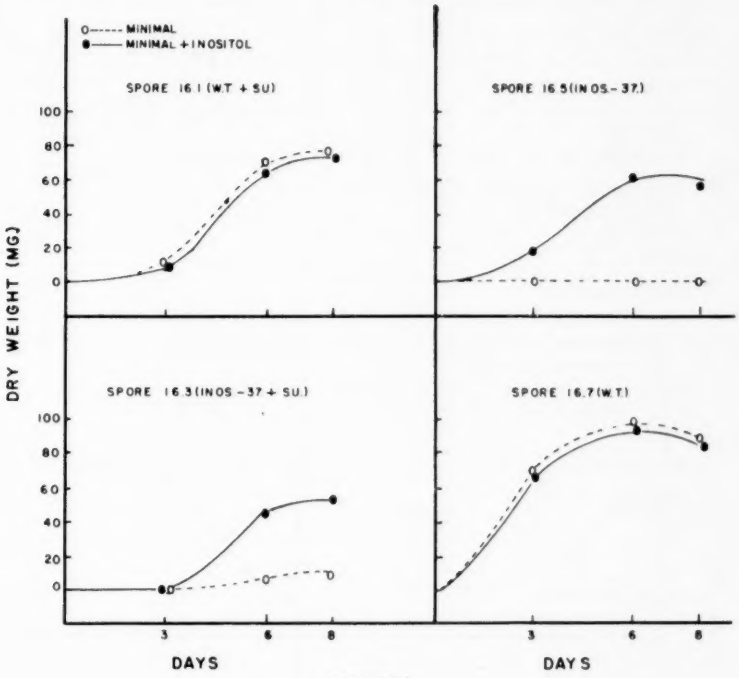


FIGURE 1

Figures 1–3 show comparative growth curves obtained in liquid minimal and minimal supplemented with 4.0 μg. inositol per ml. (20 ml. in 125-ml. Erlenmeyer flasks) for cultures from complete asci from a cross of the suppressor of inositolless 37401 (inos.-37 + Su.) with wild type (W. T.). Segregations of inositol-independent: inositolless cultures are as follows: figure 1, 6:2; figure 2, 8:0; figure 3, 4:4.

the mycelium of a suppressor strain was harvested (total dry weight was 74 mg.) and both filtrate and mycelium were tested for inositol activity. No activity was present in the filtrate, but the following values were obtained in micrograms of inositol per mg. of dried mycelium in duplicate sample tests: with the Neurospora assay, 0.19 and 0.13; with the Ashbya assay, 0.17 and 0.24. The value obtained for wild type mycelium in com-

TABLE 2

TESTS FOR SUPPRESSION OF OTHER INOSITOLLESS MUTANTS BY A SUPPRESSOR OF INOSITOLLESS 37401 (37401+Su.)

CROSSES TO INOSITOLLESS MUTANTS	NUMBER OF COMPLETE ASCI WITH SEGREGATION TYPES INOSITOL-INDEPENDENT:INOSITOLLESS 4:4	6:2	8:0
(*A*) Wild type + suppressor, culture No. 1, crossed with:			
37102	3	0	0
46316	18	0	0
46802[a]	5	0	0
64001	19	0	0
89601	21	0	0
37401	2	15	2
J.H. 2626	7	0	0
J.H. 5202	2	0	0
(*B*) Wild type + suppressor, culture No. 2, crossed with:			
89601	9	0	0
37401	1	10	0

[a] 8-Spored asci only.

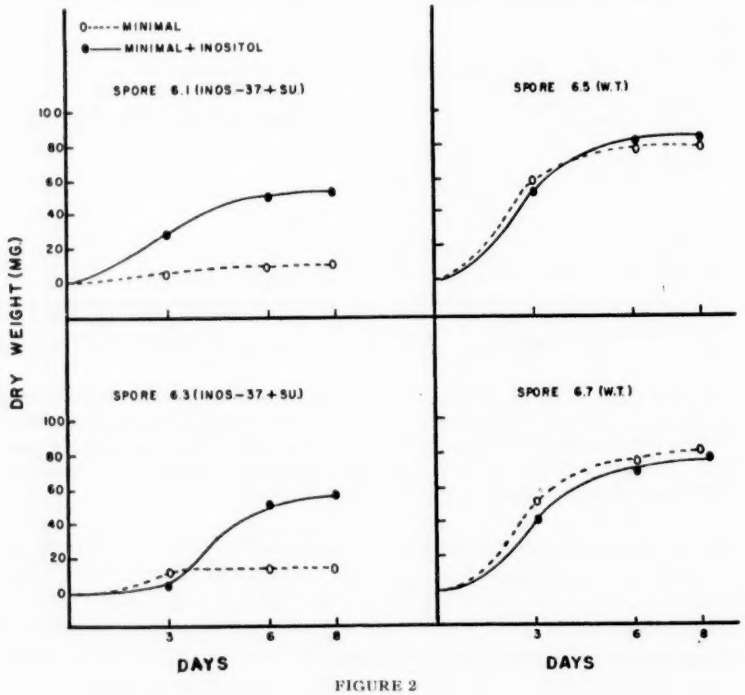


FIGURE 2

parable assays is approximately 2.0, indicating that the suppressor has only about one-tenth the mycelial concentration of inositol present in the wild type.

Genetic Tests of the Suppressor in Combination with Other Inositolless Mutants.—Following the demonstration that the reversion in 37401 was actually the result of a suppressor mutation at an independent locus, it became of interest to determine whether this suppressor would also suppress other inositolless mutants. Consequently, simultaneous crosses were made between the suppressor combined with wild type (W.T. +Su.)

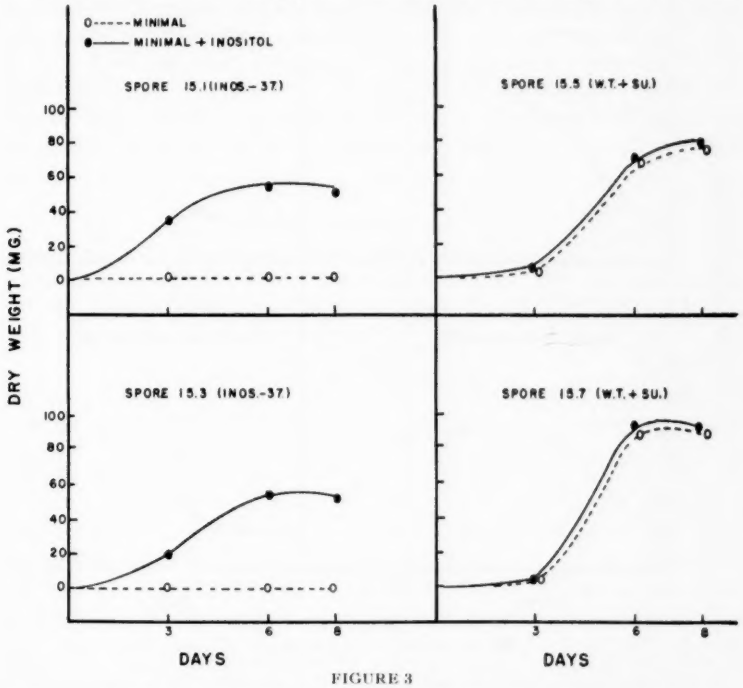


FIGURE 3

and several different inositolless mutants. Serial ascospore isolations were then performed to test for the presence of 6:2 and 8:0 inositol-independent: inositolless segregations, since more than four inositol-independent cultures in a single ascus would indicate a suppression of the particular inositolless mutant being tested by the suppressor of mutant 37401. Some of these crosses were also segregating for albino (al_2-15300) which facilitates the visual detection of any segregation irregularities.

The results of these crosses, shown in table 2, indicate that only 4:4 segregations were obtained with seven different inositolless mutants

(although relatively few complete asci were obtained with mutants 5202 and 37102), whereas mutant 37401 again showed the expected three types. Typical growth curves obtained from the four cultures in one ascus (No. 3) from a cross with mutant 89601 are shown in figure 4.

In order to be certain that the suppressor was actually present, even though inactive, with the various inositolless mutants tested, further

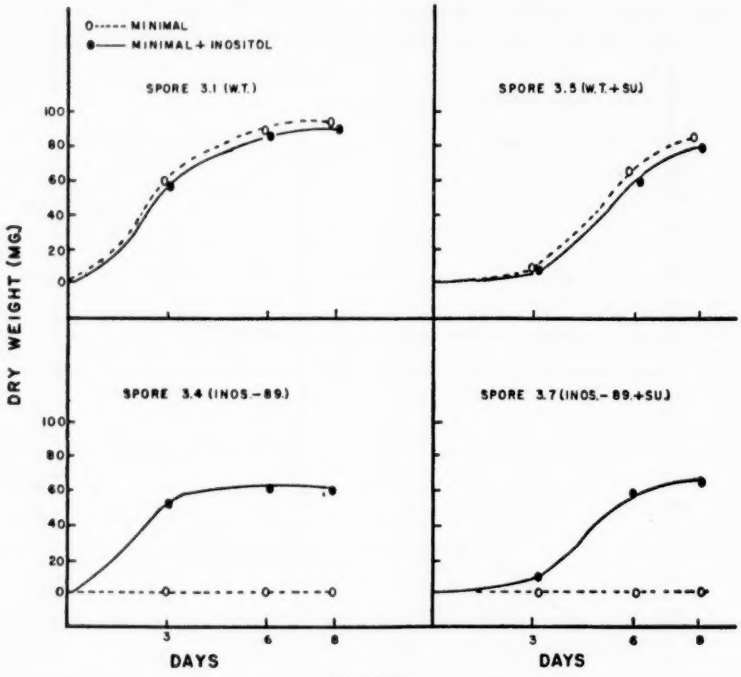


FIGURE 4

Comparative growth curves obtained by methods indicated in figures 1–3, but from a cross of the suppressor of 37401 combined with wild type (W.T. + Su.) with inositolless 89601 (inos. − 89). Cf. table 3.

crosses to 37401 were made utilizing the four cultures from ascus 3 in the cross with 89601. The results of these crosses are shown in table 3. It can be seen from the ascus segregation patterns that the suppressor can be re-extracted from mutant 89601 and is again able to suppress mutant 37401, although it did not effect growth in the absence of inositol when combined with mutant 89601.

Discussion.—The particular reversion selected for study in the present

investigation was obtained as a phenotypically distinctive, slow-growing mutant from inositolless microconidia (of mutant strain 37401) exposed to ultraviolet irradiation and plated on minimal medium. Extensive genetic analyses of phenotypically "normal" reversion types—those exhibiting a morphology and growth rate on minimal medium similar to inositol-independent, "wild type" strains[2]—have demonstrated that the vast majority of these behave in crosses as instances of reverse mutation. Preliminary evidence that certain of the slow-growing types represent instances of suppressor mutation rather than of reverse mutation was obtained when crosses of certain of these mutants to wild type yielded inositolless cultures from random ascospore isolations.[2] The present investigation gives detailed genetic evidence from serial ascospore isolations for one such mutant and demonstrates conclusively that reversions may be obtained which are able to grow in the absence of inositol as a result of suppressor mutation at a locus completely independent of the original inositolless locus.

TABLE 3

TESTS TO DEMONSTRATE THE PRESENCE OF THE SUPPRESSOR OF 37401 IN COMBINATION WITH INOSITOLLESS MUTANT 89601. CROSSES OF ONE CULTURE FROM EACH SPORE PAIR IN A COMPLETE ASCUS ISOLATED FROM A CROSS OF WILD TYPE PLUS SUPPRESSOR WITH 89601-al_2 (CF. FIG. 4)

SPORE NO.	CROSSED WITH	NO. OF COMPLETE ASCI OF SEGREGATION TYPES INOSITOL-INDEPENDENT : INOSITOLLESS 4:4	0:8	2:6	6:2	AL_2 SEGREGATION	INDICATED GENOTYPE OF SPORE
3.1*a* al_2^+	37401-al_2	15	0	0	0	1:1	W.T.
3.4*a* al_2^+	37401-al_2	0	17	0	0	1:1	inos. − 89.
3.5*A* al_2	37401-al_2^+	4	0	0	11	1:1	W.T. + Su.
3.7*A* al_2	37401-al_2^+	2	4	10	0	1:1	inos. − 89. + Su

Growth tests of the present suppressor indicate that this mutation only partially suppresses the effect of the original mutation giving an inositolless phenotype. Although growth occurs in the absence of this substance, it is much less rapid than in inositol-independent strains (either wild type or typical reverse mutation types), and growth is markedly stimulated by the addition of inositol to the medium. The possibility that some change has occurred as a result of the suppressor mutation such that inositol is no longer required by the reversion, or is being replaced by some other substance, appears to be eliminated by the demonstration by microbiological assay that *meso*-inositol is present in the mycelium of the suppressor strain, although in a reduced amount compared with typical inositol-independent types. This latter condition presumably explains the relatively restricted growth of the suppressor on minimal and its stimulation by added inositol.

Perhaps the most notable finding in the present study is the apparent marked specificity of this suppressor strain. Crosses of the suppressor in combination with the wild type allele at the inositolless locus were made to seven allelic inositolless mutants (all of independent origin). A simultaneous cross was made to mutant 37401, in which the suppressor was obtained, as a control. The expected 6:2 and 8:0 inositol-independent: inositolless segregations were recovered in serial ascospore isolations from the control cross, but in all the other crosses only 4:4 segregations were observed. Appropriate additional genetic tests with one of the mutants (89601), involving re-extraction of the suppressor, demonstrated that the suppressor was definitely present, but that no growth on minimal occurred even in protracted growth tests in tubes and in flasks.

Previous studies of the same series of inositolless mutants used in the present tests for suppressor activity have demonstrated that these mutants are all allelic (or in some instances possibly pseudoallelic), but that several differ markedly in at least one respect—the frequency with which they undergo reverse mutation spontaneously or following ultraviolet radiation.[2] With respect to its frequency of reverse mutation induced by ultraviolet, mutant 37401 appears to be distinct from the other seven inositolless types, and the present evidence for the specificity of its suppressor adds another criterion of distinction. The evidence obtained from this suppressor does not, however, aid in elucidating the problem of possible pseudoallelism at this locus. The best evidence to date for recombination is that between mutants 37401 and 64001, from crosses of which inositol-independent cultures (which have been shown not to be pseudo-wild types[6]) are regularly recovered.[2] On the other hand, no evidence of recombination has been obtained from crosses of 37401 and 37102, involving tests of comparable numbers of ascospores.[2] With both mutants, however, there is no effect of the suppressor of 37401.

A similar specificity of suppressor effect in *Neurospora crassa* for one of two apparently allelic mutants of independent origin has been reported by Yanofsky.[7] In this instance it was shown that the particular enzyme, tryptophan desmolase, which was absent from the original mutants, was present in the suppressed mutant, but could not be detected in the mutant which did not grow without tryptophan when carrying the suppressor. An additional noteworthy similarity exists between the results reported for the tryptophanless mutant and those obtained in the present study. When the suppressor mutation is combined with the wild type allele at the inositolless locus, a marked interaction occurs such that the growth of the resulting strain is greatly retarded compared with the normal wild type. Further, this effect cannot be overcome by added inositol. The nature of this interaction is not clear. At present it does not appear particularly profitable to indulge in speculations as to the biochemical mechanism by

which the present suppressor is able to overcome, at least partially, the effect of mutation at the original inositolless locus. Certain aspects of this general problem have been considered for the tryptophanless suppressor,[7] a situation in which the biochemical evidence is much more extensive than in the present case.

The present instance of a suppressor which is effective in combination with only one of a series of allelic (or in some instances possibly pseudoallelic) mutants, provides additional evidence that similarity of response by two or more mutants in Neurospora to a particular suppressor cannot be used as a criterion of allelism. Although the original investigation of suppressors in Neurospora[4] suggested that such mutants might be useful in tests for allelism, it is now apparent that the effects of suppressor mutation may be more complex than was previously anticipated.[8] In addition to instances such as the present case, in which allelic mutants are not all affected by the same suppressor, instances are also known in Neurospora in which clearly non-allelic mutants are suppressed by the same suppressor.[2, 8, 9]

Other instances of reversions are known in Neurospora inositolless mutants which behave in preliminary tests like suppressors, and at least some of those in strain 37401 appear to be different from the present mutant. Additional investigations are contemplated to determine if genetically distinct suppressors do occur. Such tests may make it possible to establish other criteria for distinguishing among the apparent isoalleles at the inositolless locus, much as has been done recently at the vermillion locus in *Drosophila melanogaster*.[10]

Summary.—A genetic analysis has been made of a slow-growing reversion obtained from platings on minimal medium of ultraviolet-treated microconidia of inositolless mutant 37401 in *Neurospora crassa*. These tests indicate that the reversion arose as a result of an independent suppressor mutation rather than by the more usual reverse mutation. The resulting suppression of the inositolless phenotype is only partial, however, since the suppressor grows much less in the absence of inositol than do typical inositol-independent strains and is markedly stimulated by the addition of inositol. Microbiological assays indicate that inositol is present in the mycelium of the suppressor but in reduced amounts compared with wild type strains. Genetic tests have been made to determine the effect of the suppressor of mutant 37401 on seven other allelic (or possibly pseudoallelic) inositolless mutants of independent origin. The results indicate that the suppressor is markedly specific for mutant 37401, since growth does not occur in the absence of inositol when the suppressor is combined with any of the other mutants.

Acknowledgment.—The authors desire to express their thanks to Mrs.

Norman Giles and to Miss Mary Case for their assistance in this investigation.

* This research was supported in part by a research contract with the Atomic Energy Commission.

[1] Giles, N. H., Jr., and Lederberg, E., *Amer. Jour. Bot.*, **35**, 150–157 (1948).

[2] Giles, N. H., Jr., *Cold Spr. Harb. Symp. Quant. Biol.*, **16**, 283–313 (1951). Also unpublished material.

[3] Partridge, C. W. H., and Giles, N. H., unpublished.

[4] Houlahan, Mary B., and Mitchell, H. K., PROC. NAT. ACAD. SCI., **33**, 223–229 (1947).

[5] Schöpfer, W. H., *Bull. Soc. Chimie biol.*, **33**, 1113–1146 (1951).

[6] Mitchell, M. B., Pittenger, T. H., and Mitchell, H. K., PROC. NAT. ACAD. SCI., **38**, 569–580 (1952).

[7] Yanofsky, C., *Ibid.*, **38**, 215–226 (1952).

[8] Mitchell, Mary B., and Mitchell, H. K., *Ibid.*, **38**, 205–214 (1952).

[9] Lein, J., and Lein, P. S., *Ibid.*, **38**, 44–48 (1952).

[10] Green, M. M., *Ibid.*, **38**, 300–305 (1952).

THE EFFECT OF PRETREATMENT WITH INFRA-RED RADIATION ON THE X-RAY INDUCED SEX-LINKED RECESSIVE LETHAL AND VISIBLE MUTATION RATE IN DROSOPHILA MELANOGASTER

BY WILLIAM J. YOUNG,* HENRY T. YOST, JR., P. T. IVES, AND R. PAUL LEVINE†

BIOLOGICAL LABORATORY, AMHERST COLLEGE, AMHERST, MASSACHUSETTS

Communicated by A. F. Blakeslee, April 16, 1953

Pretreatment with infra-red radiation has been shown to cause an increase in the number of gross chromosomal aberrations induced by x-radiation in Drosophila;[1] however, no increase in the number of recessive lethals has been found with similar treatment.[2] These findings are of interest, since recessive lethals have been thought to arise, in large part, as a direct consequence of chromosome breakage.[3–5] Kaufmann and Gay[2] have explained these results by proposing that rearrangements do not give rise to any large portion of recessive lethals, and that the infra-red acts to sensitize the breakage ends of the chromosome so that recombination results rather than restitution. Thus there is an increase in the number of gross chromosomal rearrangements without an increase in the total number of primary breaks, and, consequently, there is no increase in the number of recessive lethals. The data of Kaufmann and Gay are in general agreement with the hypothesis of Lea and Catcheside.[4]

Difficulties of interpretation arise when the foregoing arguments are

considered in the light of the data obtained with Tradescantia, using combined infra-red and x-radiation. It has been shown that all types of chromosome aberration are increased by infra-red pretreatment and that the total number of primary breaks is increased.[6-8] If the mechanism of action of infra-red be the same in both forms, and if the recessive lethals be directly related to chromosome breakage, there should be an increase in recessive lethals with infra-red pretreatment. Since this discrepancy existed, it was felt that further investigation of the recessive lethal problem would be worth while.

Recently two factors which modify the action of infra-red radiation have been studied. It has been shown that the temperature at the time of radiation is critical,[6] and that the intensity of infra-red radiation may be of importance.[7] Swanson (personal communication) has demonstrated that the most effective temperature for Tradescantia is about 12°C., and the maximum temperature is about 22°C. It was felt that the effective temperature for Drosophila might be lower than that previously tried (18°C.)[1] and that an increase in intensity of infra-red might influence the rate of recessive lethal induction. Experiments were undertaken to test these points.

At the same time, a weakness in the test for recessive lethals was apparent. The induction of more than one lethal on the X-chromosome would be missed. Since the x-ray and infra-red are additive, and since there is usually found a large increase in the number of rearrangements, it seemed likely that a chromosome which had been "hit" by x-radiation (that is, had one primary break) would be likely to have more than one break if previously treated with infra-red radiation. For this reason attempts were made to determine the number of lethals on some of the tested X-chromosomes by studies of crossing over. Further, it was hoped that the use of visible mutations might help to elucidate the problem of multiple events and at the same time give an independent estimate of the induced mutations.

Experimental Procedures and Results.—A strain of Oregon-R, inbred for 70 generations to insure maximum homozygosis[8] was used as the experimental stock. During the course of the experiment the stock was kept in bottle cultures without selection or inbreeding. Males to be treated were selected between two and seven days after eclosion, to minimize the variation in the natural mutation rate.[9] The males were placed in shell vials and allowed to equilibrate to the cold-room temperature (15°C.) before immersion in the water bath, which averaged 8°C. At this temperature the flies sit motionless on the food so that by slanting it all the flies are exposed quite readily to the radiation. The vials were immersed in the water bath during the infra-red treatment. Control vials were placed in the same bath, protected from the radiation by a black rubber covering. The tem-

perature of the water bath at the end proximal to the radiation source was found to increase less than 0.5°C. over that at the control end, 37 cm. distant. The exposed vials were held within 5.5 cm. of the Corning No. 2540 filter, through which all radiation passed. The infra-red source was a General Electric 250-watt heat lamp; the distance from the source to the treated flies was about 13 cm. This distance is less than that used by Kaufman, *et al.*,[1] and therefore the infra-red radiation was given at a greater intensity. Both control and infra-red treated flies were removed from the

TABLE 1

CONTROL DATA: NO X-RAY TREATMENT

TREATMENT	TESTS	LETHALS	SEMI-LETHALS	VISIBLES	TOTALS	%	P
4 hrs. IR plus 8°C.	588	1	0	0	1	0.17	0.5–0.6
8°C. alone	243	0	0	0	0	0.00	
4 hrs. IR plus 8°C.	339	2	0	0	2	0.59	0.625
8°C. alone	313	1	1	0	2	0.64	
4 hrs. IR plus 8°C.	566	0	0	0	0	0.00	
8°C. alone	537	0	0	0	0	0.00	

TABLE 2

SUMMARY OF DATA: COMBINED X-RAY AND INFRA-RED TREATMENTS

GROUP NO.	TREATMENT	TESTS	LETHALS	SEMI-LETHALS	VISIBLES	TOTALS	%	P
1	2500 r	423	23	4	4	31	7.3	
	4 hrs. IR plus 2500 r	479	23	10	1	34	7.1	0.7–0.8
2	2500 r	453	15	3	3	21	4.6	
	4 hrs. IR plus 2500 r	472	31	4	5	40	8.5	0.02
3	2500 r	537	28	2	6	36	6.7	
	4 hrs. IR plus 2500 r	516	20	5	2	27	5.2	0.3–0.5
4	2500 r	468	30	5	6	41	8.8	
	4 hrs. IR plus 2500 r	515	39	2	3	44	7.5	0.90–0.95
5	2500 r	522	32	3	1	36	6.9	
	4 hrs. IR plus 2500 r	522	24	4	6	34	6.5	0.8–0.9
6	1800 r	586	29	4	6	39	6.7	
	5 hrs. IR plus 1800 r	587	16	3	6	25	4.3	0.05–0.10
7[a]	1000 r	956	...	...	...	26	2.7	
	7 hrs. IR plus 1000 r	936	...	...	...	37	4.0	0.1–0.2

[a] Lethals not retested.

water bath and x-rayed at the same time. All x-ray exposures were made with a General Electric industrial machine, operated at 80 kv. and 10 ma.[10] The dosages for both types of radiation varied between experiments and are given with the data.

Tests for induced mutations were carried out with the Muller-5 stock.[11] Any vial in which three or more wild type males could be seen simultaneously was classed as non-mutant; that is, no recessive lethal or visible mutation had been produced in the chromosome tested. All vials having less

than three wild males, or whose males exhibited a visible mutation, were retained. The field of the microscope showed a maximum of thirty flies, and the presence of three wild males, indicating an excess of 10 per cent wild males, was considered not to be a lethal or semi-lethal mutation. All visible and lethal mutants were retested. Any case giving a total proportion of more than 10 per cent of wild males in not less than one hundred counted flies in the two generations was considered to be normal. Those counts giving under 10 per cent and over 0 per cent were scored as semilethals. In all cases, control and experimental series were handled simultaneously, with the vials randomized in such a way that the scorer was unaware whether he was examining experimental or control vials. Crossover tests were made of a selected group of lethals, using the *y ct*^{6} *ras*^{2} *f* stock. Comparison of data from different series was made by means of the Chi-square test.

Table 1 presents the data derived from infra-red and cold treatments. It is evident that cold shocks alone, or in combination with infra-red treatments, are not capable of increasing the frequency of sex-linked recessive

TABLE 3

COMBINED ANALYSIS OF THE DATA PRESENTED IN TABLE 2

GROUP	TESTS	TOTAL MUTANTS	MUTATION, %	*P*
All controls (2500 r)	2403	165	6.9	0.18
All combined treatments (4 hrs. IR plus 2500 r)	2504	179	7.2	0.21
Both groups combined (line 1 plus line 2)	4907	344	7.0	0.22

lethals or visibles. Chi-square tests of the combined six series gave a *P* value of 0.1–0.2. The average mutation rate for all groups combined (0.19%) did not differ from that obtained by Kaufmann and Gay[2] with 48 hours of infra-red treatment in the Swedish-b stock (0.17%).

Table 2 presents the results of runs in which combined infra-red and x-radiation was given. Groups 1 through 5 are replications of one treatment. Groups 6 and 7 are modifications of this treatment. Only in group 2 was there an apparent difference in the frequency of lethals recovered. In this case, the data suggested that pretreatment definitely increased the recessive lethal mutation rate. However, as table 3 establishes, the difference observed is best considered to be simply a chance deviation from the mean. This conclusion is based on the Chi-square values obtained by: (*a*) lumping all control treatments, (*b*) lumping all experimental treatments, and (*c*) lumping all ten series. In each case the *P* value is of the order of 0.20, well above the level of significance. Therefore, it must be concluded that pretreatment with infra-red, under these conditions, does not give a significant increase in the number of x-ray induced sex-linked lethal and

visible mutations obtained. The data from groups 6 and 7 indicate that this conclusion is valid at lower x-ray doses.

A group of lethals was selected from the combined treatment series (group 5, table 2) to be tested by crossing over studies, as it was felt that the effect of the infra-red might be obscured by the presence of more than one lethal on the X-chromosome. Table 4 presents the data derived from these tests. Since only one stock out of the 40 tested showed the presence of more than one lethal, it must be concluded that the combined treatment did not produce a significantly greater number of chromosomes with more than one lethal than would be expected.

Discussion.—The experiments presented in this paper were undertaken to determine whether an increase in recessive lethals can be obtained with infra-red pretreatment when the modifying factors are suitably controlled. Data presented by Swanson and Yost[6] have shown that the infra-red effect is dependent upon the temperature at which the infra-red is given, with low temperature enhancing the effect. However, the data presented here

TABLE 4

DISTRIBUTION OF LETHALS

TYPE	TOTALS
Single lethals	33
Lethals with inversions	4
Lethals with deficiencies	2
Double lethals	1
TOTAL	40

show quite clearly that lowering the temperature does not result in an increase in recessive lethals. When this was discovered, it was thought that perhaps the effect was being masked by the induction of double lethals. This would be likely to happen if the effect of the infra-red were dependent upon the summation of infra-red and x-ray energy. Only the chromosome which received x-ray energy could be affected, and that one might be severely damaged. To test this possibility, some of the lethals derived from the combined radiation series were tested for the presence of more than one lethal by crossing over studies. As can be seen in table 4, the results were negative. It is true that the markers are far enough apart to permit two closely linked lethals to appear as one in some cases. I may be assumed, however, that in most cases any two lethals would be far enough apart to be detected by the technic used, so long as the induction is random. It is possible to postulate that double lethals arising from combined treatments may be non-random; that is, a particular x-ray track might produce two closely placed lethals in a pretreated chromosome. Any lethals which lie closer together than five map units would go unde-

tected by crossover studies. Such an hypothesis is essentially untestable, however, and should be disregarded. From the foregoing, it is evident that there is no increase in sex-linked recessive lethals when infra-red is used as a pretreatment to x-radiation. These results are in accord with those obtained by Kaufmann and Gay.[2] Similarly, there is no increase in visible mutations.

In discussing their data, Kaufmann and Gay suggest that there is little role of position effect in the production of recessive lethals. This follows from the finding that there is an increase in gross chromosomal rearrangement with infra-red pretreatment without an increase in recessive lethals. Such findings are in agreement with the hypothesis of Lea and Catcheside[4] on the origin of recessive lethals. On the other hand, Kaufmann and Gay state that, on the basis of this hypothesis, one would expect a decrease of the order of 5 per cent in recessive lethals with supplementary treatment, and their data indicate such a decrease. The data presented here do not show any such decrease. If there is any difference (and there is none shown by statistical methods), it is in the direction of a slight over-all increase. Nevertheless, since it is evident that there is no induced increase of recessive lethals, and since it has been shown that gross chromosomal aberrations are increased,[1] it is possible to state that there is little apparent relationship between the two.

Since the publication of the earlier work on Drosophila,[2] evidence has been presented which indicates that in Tradescantia the action of infra-red is to increase the number of primary breaks, not to modify the method of recombination.[6-8] Kaufmann and Gay start with the assumption that the recessive lethals induced in Drosophila originate with chromosomal breakage in large part. Having made such an assumption it is necessary to assume further that the action of infra-red is to modify the recombination process and not to increase the number of primary breaks; for if there were an increase in breakage *per se*, it would result in an increase in the number of recessive lethals. The basic assumption about the origin of lethals is not necessary, however. In fact, one might turn this argument around and assume that if there is no increase in recessive lethals with infra-red pretreatment, there is no causal relationship between breakage and the formation of recessive lethals. Without this latter argument, it is necessary to propose two quite different modes of action of the infra-red in Tradescantia and Drosophila. Such an hypothesis seems extreme even though it is clear that extrapolation from one form to another is dangerous with distantly related types. Furthermore, there is some evidence from natural populations of Drosophila which bears on this problem; that is, the fact that recessive lethals have a high natural mutation rate compared to other types of mutations. If we are to assume that chromosome breakage is necessary for the induction of a large portion of recessive lethals, we should

find a high degree of correlation between recessive lethals and chromosomal rearrangement in natural populations. Ives,[12] in a survey of one local wild population, has shown that while 7.5 per cent of 159 lethal chromosomes carried inversions, 116 non-lethal chromosomes from the same population had an inversion frequency of 8.6 per cent. Clearly there is no relation in those data between chromosomal breakage and recessive lethal mutations. Further studies of this kind (by Levine and Ives) are in progress in this laboratory, utilizing artificial cage populations of many thousands of individuals related to Oregon-R and to the local wild population.

The question which arises in this connection is whether we can logically assume that there is a higher frequency of chromosome breakage than of point gene mutation. (Point gene mutation is here interpreted as a chemical change not involving the entire chromosome strand.) This seems unlikely. It has been demonstrated quite clearly that a high proportion of x-ray induced recessive lethals occur near breakage points in the chromosome;[3] but this is not evidence that there is a causal relationship. It is perfectly evident that the breaks give us visible proof that some change has taken place; but the order of magnitude is so great in such changes that it is not unlikely that mutations should take place in the same vicinity. If we consider the mutation process in the light of the model proposed by McElroy and Swanson,[13] it is clear that the rate determining factor is the energy of activation of the chemical change which results in mutation. Obviously different processes will have different energies of activation. Unless we assume that *all* mutation is accompanied by chromosomal breakage (in the usual sense), we can see that when the x-ray provides the energy necessary to accomplish chromosome breakage, it must at the same time provide energy which surpasses the energy of activation of other mutation processes. The question which arises at this point is why the combined infra-red and x-ray energy does not promote mutation. There are two factors to be considered regarding this problem. Since the absorption of the energy of each radiation is separate in time, and since neither provides sufficient energy singly to cause chromosome breakage (here considering the "increased" portion of the breakage), it is possible that the total energy incorporated in the chromosome at any one time is not sufficient to cause both chromosome breakage and mutation. Secondly, the infra-red effect depends upon the absorption of specific wave-lengths of infra-red, and consequently only certain regions of the chromosome will be sensitized by pretreatment. When the infra-red energy is absorbed, these regions are effectively weakened, or made more susceptible to breakage by x-ray energy.[6] With the addition of energy derived from x-radiation, these regions, predisposed to breakage, undergo disunion. The pretreatment with infra-red has determined both the quantity and the quality of the reaction in this manner; there is no excess energy to be utilized by re-

lated structures for mutation to any appreciable extent. Regardless of the mechanism, it seems permissible to conclude from the data presented that the recessive lethal is not dependent upon chromosomal breakage. Such conclusion is borne out by the lack of effect of infra-red pretreatment on the induction of visible mutations.

Summary.—Pretreatment with infra-red radiation does not increase the number of sex-linked recessive lethal or visible mutations induced by x-radiation, even though the environmental factors known to influence the action of infra-red are favorable. Crossing over studies of the lethals failed to indicate any significant portion of the chromosomes with two or more lethals resulting from the combined radiations. It is concluded from these and other findings and from theoretical considerations that recessive lethal and visible mutations do not result from chromosome breakage and that neither type of mutation is to be expected as a result of infra-red radiation acting alone or in pretreatment before x-radiation.

* Present address: Department of Biology, The Johns Hopkins University, Baltimore, Md.

† The organization and scoring of the experiments was done under the supervision of Ives. The radiation was carried out by Yost. The routine of the experiments was shared by all authors but carried chiefly by Young. The manuscript was prepared chiefly by Young and Yost and is adapted in part from Young's thesis for the degree of Master of Arts at Amherst College (1952).

[1] Kaufmann, B. P., Hollaender, A., and Gay, H., *Genetics*, **31**, 349–367 (1946).

[2] Kaufmann, B. P., and Gay, H., these PROCEEDINGS, **33**, 366–372 (1947).

[3] Demerec, M., *Cytologia*, Fujii Jubilee Volume, 1937, pp. 1125–1132.

[4] Lea, D. E., and Catcheside, D. G. J., *Genetics*, **47**, 10–24 (1946).

[5] Herskowitz, I. H., *Ibid.*, **36**, 356–363 (1951).

[6] Swanson, C. P., and Yost, H. T., these PROCEEDINGS, **37**, 796–802 (1951).

[7] Yost, H. T., *Genetics*, **37**, 457–468 (1952).

[8] Ives, P. T., *Ibid.*, **30**, 167–196 (1945).

[9] Ives, P. T., *Evolution*, **4**, 236–252 (1950).

[10] The authors are indebted to Profs. Carl Kaiser and Karl Hendrickson of the University of Massachusetts, who made the x-radiation facilities available for this study.

[11] Spencer, W. P., and Stern, C., *Genetics*, **33**, 43–74 (1948).

[12] Ives, P. T., *Evolution*, **1**, 42–47 (1947).

[13] McElroy, W. D., and Swanson, C. P., *Quart. Rev. Biol.*, **26**, 348–363 (1951).

A THEORY OF N-PERSON GAMES WITH PERFECT INFORMATION*

BY DAVID GALE

BROWN UNIVERSITY

Communicated by J. von Neumann, April 17, 1953

1. Introduction.—The theory of the general n-person game, in contrast to that of the zero-sum two-person game, remains in an unsettled state. The chief problem seems to be that of determining the proper definition of a solution for such games. The efforts in this direction divide themselves into two groups, the cooperative theory in which the players are expected to form coalitions, and the non-cooperative in which such coalitions are forbidden. The first group includes the theory of von Neumann and Morgenstern[1] and the more recent work of Shapley,[2] the second the equilibrium point theory of Nash.[3] The cooperative theories have the shortcoming that the notion of solution does not give much insight into how the games should be played. The non-cooperative theory is more specific in this regard, but unfortunately the definition of solution only applies to particular games and as yet no way has been found to determine *a priori* whether a given game is solvable in the non-cooperative sense. Certainly many interesting games are not solvable by this definition.

The theory to be outlined below gives a very specific notion of solution for a special but important class of n-person games, namely, those with perfect information. The theory is non-cooperative in the sense described above. The most general theorem in the literature concerning such games is due to Kuhn[4] and states that all such games possess an equilibrium point among their pure strategies. This equilibrium point will not in general be unique, nor will such games be solvable in the sense of Nash. We therefore approach the problem from a different point of view. Our definition of solution leads to a particular set of strategies which form an equilibrium point, but these will generally be mixed rather than pure strategies.

2. Some Examples.—An example will be useful in motivating the subsequent definitions.[5] An integer between 1 and 9 is to be chosen at random by an umpire. Three players are to choose in turn distinct numbers from this range, and the person guessing closest to the chosen number receives one dollar. In case of a tie each player wins a half dollar. A very naive first player might guess 5 initially. A little further thought reveals, however, that the second and third players would be expected to choose numbers 4 and 6 so that the first player's expectation would be $^1/_{10}$ as compared to $^4/_{10}$ for players 2 and 3. A "correct" way to play the game would be for player I to choose either 3 or 7, player II, 7 or 3,

and player III any remaining number. A rough justification for these choices is the following. If players I and II have already chosen, the choice for player III is clear, he will choose the number which maximizes his expectation. If there are several numbers which do this, we assume he is equally likely to pick any one of them. This amounts to the assumption that player III is only interested in maximizing his own expectation and is not concerned at all with the expectations of the others. Now assume player I has chosen. Player II tentatively considers each possible choice under the assumption that player III on the next move will play in accordance with the principles already described. He makes that choice which leaves him with the largest expectation under these assumptions. Now consider the initial choice for player I. For each possible choice he can compute how players II and III will choose in accordance with the principles described above. Thus, in each case he computes his expectation and plays so as to maximize it. Again, if there are several choices which maximize, he chooses among them by a random device. The expectations for players I, II, and III if they play as described will be $^{7}/_{18}$, $^{7}/_{18}$, and $^{4}/_{18}$.

The theory to be outlined here makes precise the notions described above for any game with perfect information and provides a definition of values and optimal strategies for such games.

The reader may wish to amuse himself by solving in the sense of the above example the following continuous version. Three players choose in turn a real number on the unit interval. A player is said to "control" the number $t \in [0, 1]$ if t is closer to his choice than to that of the other players. At the end of the game each player is paid an amount equal to the length of the interval he controls.

3. Definitions and Notation.—We shall consider here only finite games, and for the sake of clarity and keeping notation under control, we shall assume that the games involved do not have chance moves although all our results carry over to this case.

The notations used here are essentially those introduced in an earlier paper.[6] We summarize them briefly.

Definition: An *n-person game* Γ with perfect information consists of:

(1) A set X called the set of *positions* of Γ.

(2) A point $x_0 \in X$ called the *initial position*.

(3) A function φ called the *predecessor function* which maps $X - x_0$ into X and has the property: (A) There is an integer N such that $\varphi^N(X) = X_0$. (Intuitively, the positions which can be reached on the next move from positions x are the positions $\varphi^{-1}(x)$.) The set of positions x such that $\varphi^{-1}(x) = \emptyset$ is called the *set of final positions* and is denoted by X_F.

(4) A partition of $X - X_F$ into sets $(X_1, X_2 \ldots X_n)$. (The position $x \in X_i$ if the ith player is to move at position x.)

(5) n real valued functions $P_1, P_2, \ldots, P_n$ defined on the set X_F. P_i is called the ith *player's pay-off function.*

The above five objects constitute the game Γ. It remains to describe how the game is to be played. The intuitive notion is the following: Suppose $x_0 \in X_i$. Then player i chooses a position, $x_1 \in \mathcal{P}^{-1}(x_0)$. Now suppose $x_1 \in X_{i_1}$. Then player i_1 chooses a position $x_2 \in \mathcal{P}^{-1}(x_1)$, etc. Play continues in this way until a position $x \in X_F$ has been chosen, which will always occur because of property (A). At this point the play has terminated and player i receives the amount $P_i(x)$, $i = 1, \ldots n$.

As is customary in game theory (but not in practice!), we assume that each player has chosen his "strategy" in advance, that is, he has made up his mind as to what choice he will make at all conceivable positions. The following definitions make these notions precise.

Definitions: A *play* of Γ is a sequence $(x_0, x_1, x_2, \ldots, x_k)$ where $\mathcal{P}(x_i) = x_{i-1}$, and $x_k \in X_F$.

A *pure strategy* σ_i for player i is a function from X_i into X such that for each $x \in X_i$, $\sigma_i(x) \in \mathcal{P}^{-1}(x)$.

Given an n-tuple of pure strategies $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ we define inductively a play $(x_0, x_1, \ldots, x_k)$ as follows: x_0 is the initial position. Suppose x_j has been defined and $x_j \in X_i$. Then $x_{j+1} = \sigma_i(x_j)$. It follows from the properties of the σ_i that the sequence so defined is a play, which we call the *play determined by* $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and denote by the symbol $<\sigma_1, \sigma_2, \ldots, \sigma_n>$. Finally, if x_k is the final position of the play $<\sigma_1, \sigma_2, \ldots, \sigma_n>_2$ we define $<\sigma_1, \sigma_2, \ldots, \sigma_n>_i = P_i(x_k) i = 1, 2, \ldots, n$. Thus, $<\sigma_1, \sigma_2, \ldots, \sigma_n>_i$ is the pay-off to player i if the strategies $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ are chosen by the n-players.

(There is a standard geometric picture of Γ. The positions are represented by points in a plane. If $x = \mathcal{P}(x')$ a directed line segment is drawn from x to x'. One can arrange the points x so that these lines do not intersect, thus the figure becomes a topological tree in the plane, and each play is represented by a path of segments running from the initial position x_0 to some final position. Because of its intuitive appeal, some authors have used the topological tree as a primitive concept in defining a game. From the mathematical point of view there seems little purpose in introducing these geometric irrelevancies into the theory. Further, the predecessor function seems easier to work with. The tree picture may nevertheless be useful in helping one to visualize many of the properties of games.)

For the work that follows it is convenient to make a small modification of the concept of pure strategy. In general each time a player makes a move he precludes the possibility of certain positions occurring. It would clearly be a waste of time for him to plan his strategy for positions which he knows cannot occur. Following Kuhn we call the position x irrelevant

for the strategy σ_i if x belongs to no play $<\sigma_1, \sigma_2, \ldots, \sigma_i, \ldots, \sigma_n>$. Otherwise x is called relevant, and a subset $X' \subset X$ is called relevant for σ_i if all its positions are relevant. Our modification consists in requiring that each strategy σ_i be defined only on a maximal relevant subset of X_i, relative to σ_i. This condition effects a reduction in the number of pure strategies.

Finally we recall the definition of an *n-person game in normal form*. Such a game consists of n sets $\sum_1, \sum_2, \ldots, \sum_n$ whose elements are called strategies, and n real valued functions $P_1, P_2, \ldots, P_n$ whose domain is the set $\sum_1 \times \sum_2 \ldots \times \sum_n$. If in our game Γ we let $\sum_i$ denote the set of pure strategies for player i, and define

$$P_i(\sigma_1, \sigma_2, \ldots, \sigma_n) = <\sigma_1, \sigma_2, \ldots, \sigma_n>_i,$$

we see that we have defined a game in normal form corresponding to Γ which we shall call the *normal form of* Γ.

4. The Theory.—In this section we shall define a new notion of solution of a game. Games in general will not possess solutions in our sense. However, our fundamental existence theorem states that games with perfect information always have such solutions. Our definition of solution will determine a unique optimal strategy for each player and a set of values or expectations for each player provided all play optimally.

A further preliminary remark is of interest. In examining the reasoning used to justify our solution of the example of section 2, it appears that essential use has been made of the extensive form of the game involved. In particular the order in which the players moved played a central role in the argument. It is somewhat surprising, then, that the same solution could have been obtained by examining only the normal form of the game. Indeed we shall define our notion of solution from the normal form alone. Thus, vastly different games will have the same solutions provided they have the same normal form.

Definition: The pure strategy σ_1 *dominates* σ_1', written $\sigma_1 > \sigma_1'$, provided

(i) $<\sigma_1, \sigma_2, \ldots, \sigma_n>_1 \geq <\sigma_1', \sigma_2, \ldots, \sigma_n>_1$ for all $(\sigma_2, \ldots, \sigma_n)$.

(ii) $<\sigma_1, \sigma_2, \ldots, \sigma_n>_1 > <\sigma_1', \sigma_2, \ldots, \sigma_n>_1$ for some $(\sigma_2, \ldots, \sigma_n)$.

The pure strategies σ_1 and σ_1' are *equivalent* if

$$<\sigma_1, \sigma_2, \ldots, \sigma_n>_1 = <\sigma_1', \sigma_2, \ldots, \sigma_n>_1 \text{ for all } (\sigma_2, \ldots, \sigma_n).$$

Similar definitions are made for strategies σ_i, $i = 2 \ldots n$.

Our concept of a solution is based on the following two assumptions:

(I) A player will not use a dominated strategy.

(II) A player will use equivalent strategies with equal probability.

A word concerning these assumptions. The notion of dominance has occurred frequently in the study of games, although the usual definition is not precisely the one given here. The second assumption asserts the

indifference of each player concerning the fate of the others. It might be argued that this condition is unnatural or too restrictive. For example in a two-person game a player might decide, other things being equal, to do the most possible damage to his opponent, or, if he were more amiable, to help his opponent. Actually we could develop the theory equally well by replacing assumption (II) by various other assumptions. The only requirement is that we shall have some definite rule which will determine the behavior of each player when he is confronted with equivalent strategies, that is, a rule which makes correspond to any set of equivalent pure strategies a probability combination of these strategies. Our indifference assumption, (II), is the simplest such rule, but others would serve equally well.

Corresponding to assumptions (I) and (II) we now introduce two operations on the normal form of Γ.

Operation D_1: Deleting all dominated strategies.

Operation A_1: Replacing all maximal sets of equivalent strategies by a single mixed strategy using each equivalent pure strategy with equal probability ($A \sim$ averaging).

It will be observed that these two operations, when possible, have the effect of reducing the number of strategies for each player.

We now define a new game Γ^1 in normal form whose pure strategies are those obtained from the original sets by operations D_1 and A_1. The pay-off functions P_i^1 are the original pay-off functions P_i extended in the usual way to mixed strategies whenever necessary.

We now define operations D_2 and A_2 to be similar to D_1 and A_1 but operating on the game Γ^1. We thus obtain a game Γ^2. Since the original number of pure strategies was finite, it is clear that eventually this process must terminate, i.e., no further reduction will be possible.

FUNDAMENTAL THEOREM. *If Γ is the normal form of a game with perfect information, the successive applications of the operations D_i and A_i will after a finite number of stages reduce Γ to a game Γ^k having exactly one pure strategy for each player.*

It is now clear how to define the solution of Γ. Since a pure strategy for Γ^k is a strategy (in general, mixed) for Γ, we define these unique strategies as the optimal strategies for Γ and the values are, of course, the respective pay-offs to each player when these strategies are used.

The proof of the fundamental theorem is quite involved due to the rather complicated structure of games in extensive form. The usual technique of performing an induction on the length of the game does not seem to work here. The proof proceeds by showing that at any stage of the reducing process, if any player has more than one strategy left, then further reductions are possible.

As remarked earlier, the optimal mixed strategies form an equilibrium

point. This shows then that if Γ is two-person zero-sum, our definition of value agrees with the usual one and our optimal strategies are optimal in the accepted sense. The reason we obtain a unique optimal strategy is of course due to the averaging process. This is rather unnatural in some cases; for instance, if a player knew several optimal strategies for chess he would probably use the one which would terminate the game most rapidly. Introducing such considerations would, however, complicate the theory beyond reason.

* Written under contract with the National Science Foundation.

[1] Neumann, J. von, and Morgenstern, O., *The Theory of Games and Economic Behavior*. Princeton University Press, 1947.

[2] Shapley, L. S., "A Value for n-Person Games," *Contributions to the Theory of Games—II, Annals of Mathematics*, Study No. 28, Princeton University Press, 1953.

[3] Nash, J. F., "Non-cooperative Games," *Ann. Math.*, **54**, 286–295 (1951).

[4] Kuhn, H. W., "Extensive Games," PROC. NATL. ACAD. SCI., **36**, 570–576 (1950).

[5] Examples of this type were brought to my attention by Mr. J. D. Ellinwood of Brown University, and it was from consideration of these examples that I was led to the investigations reported on herein.

[6] Gale, David, and Stewart, F. M., "Infinite Games with Perfect Information," *Contributions to the Theory of Games—II, Annals of Mathematics*, Study No. 28, Princeton University Press, 1953.

THE SPINOR CONNECTION IN THE UNIFIED EINSTEIN THEORY OF RELATIVITY

BY VÁCLAV HLAVATÝ

GRADUATE INSTITUTE FOR APPLIED MATHEMATICS, INDIANA UNIVERSITY, BLOOMINGTON, INDIANA*

Communicated by T. Y. Thomas, March 26, 1953

1. Introduction.—The Einstein unified theory of relativity is based[1] on a real tensor $g_{\lambda\mu}$ which is neither symmetric nor skew-symmetric. We denote its symmetric (skew-symmetric) part by $h_{\lambda\mu}(k_{\lambda\mu})$,[2] assume that $h_{\lambda\mu}$ is of the signature $- - - +$ and regard it as the metric tensor of the space X_4 of the unified theory of relativity. Moreover, we denote by g, h, k the determinants of $g_{\lambda\mu}$, $h_{\lambda\mu}$ and $k_{\lambda\mu}$, respectively, and assume throughout this paper

$$gh \neq 0.$$

The structure of $g_{\lambda\mu}$ depends on k as well as on

$$D \equiv \left(\frac{1}{2} hk^{\alpha\beta}k_{\alpha\beta}\right)^2 - 4hk.$$

The tensor $g_{\lambda\mu}$ being real we have only the following possibilities

$$\begin{array}{lll} (a) & D \neq 0, & k \neq 0 \\ (b) & D \neq 0, & k = 0 \\ (c) & D = 0, & k = 0. \end{array} \qquad (1)$$

Throughout this paper we confine ourselves to the cases (1a) and (1b) only. The main scope of this paper is to find a connection of the spinor space associated with X_4 *in terms of the given data* (*i.e.*, $g_{\lambda\mu}$) *only*. However, we confine ourselves to results without proofs. A more detailed study with the proofs will be published later.[3]

2. *Non-holonomic Frame in* X_4.—Denote by $((L_B{}^A))$ a diagonal matrix whose diagonal elements are

$$e^{2i\varphi},\ e^{-2i\varphi},\ e^{2\psi},\ e^{-2\psi}$$

and denote by $\overset{-1}{L}{}_B{}^A$ the elements of the inverse matrix to $((L_B{}^A))$. Moreover, let us introduce the matrices

$$((h_{AB})) = ((h^{AB})) = \begin{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \end{bmatrix}.$$

There are four linearly independent vectors $E_\lambda{}^{\mathrm{I}}, \ldots E_\lambda{}^{\mathrm{IV}}$ determined uniquely by $g_{\lambda\mu}$ up to the transformations

$$(a) \qquad {}'E_\lambda{}^A = L_B{}^A E_\lambda{}^B \qquad (2)$$

and four linearly independent contravariant null vectors

$$E_A{}^\nu = h_{AB} E_\lambda{}^B h^{\lambda\nu}$$

known up to the transformations

$$(b) \qquad {}'E_A{}^\nu = \overset{-1}{L}{}_A{}^B E_B{}^\nu \qquad (2)$$

such that

$$\begin{array}{ll} h_{AB} = h_{\lambda\mu} E_A{}^\lambda E_B{}^\mu, & h^{AB} = h^{\lambda\mu} E_\lambda{}^A E_\mu{}^B \\ E_A{}^\lambda E_\lambda{}^B = \delta_A{}^B, & E_A{}^\nu E_\lambda{}^A = \delta_\lambda{}^\nu. \end{array} \qquad (3)$$

The vectors $E_\lambda{}^{\mathrm{I}}$, $E_\lambda{}^{\mathrm{II}}$ (as well as $E_{\mathrm{I}}{}^\nu$, $E_{\mathrm{II}}{}^\lambda$) are complex conjugate, the remaining vectors are real. The functions $E_\lambda{}^A$, $E_B{}^\nu$ may be thought of as intermediate components of the unit tensor which relates the holonomic components v^ν (w_λ) of a vector and their non-holonomic components

$$v^A = E_\lambda{}^A v^\lambda \qquad w_B = E_B{}^\lambda w_\lambda$$

(so that in particular h_{AB}, h^{AB} are non-holonomic components of $h_{\lambda\mu}$ and $h^{\lambda\mu}$, respectively). The change of frame (2) changes the non-holonomic components according to

$$(c) \quad {}'v^A = L_B{}^A v^B, \qquad {}'w_B = \overset{-1}{L}{}_B{}^A w_A. \tag{2}$$

3. The Local Spinor Space $S_3(P)$.—Let P be a generic point of X_4. Denote by $T_4(P)$ its tangential space at P and by $I_3(P)$ the ideal space of $T_4(P)$. The components $v^r(P)$ $(w_\lambda(P))$ of a vector in $T_4(P)$ may be looked upon as homogeneous projective coordinates of a point (plane) in $I_3(P)$. If in particular

$$h_{\lambda\mu} v^\lambda v^\mu = 0$$

we say that v^λ is an *isotropic* point. Every other point will be termed an *anisotropic* point. Introduce the matrices associated with the $E_\lambda{}^A$

$$((E_{\lambda a}{}^b(P))) = \left[\begin{bmatrix} 0 & 0 & E_\lambda{}^{\mathrm{I}} & E_\lambda{}^{\mathrm{III}} \\ 0 & 0 & E_\lambda{}^{\mathrm{IV}} & -E_\lambda{}^{\mathrm{II}} \\ E_\lambda{}^{\mathrm{II}} & E_\lambda{}^{\mathrm{III}} & 0 & 0 \\ E_\lambda{}^{\mathrm{IV}} & -E_\lambda{}^{\mathrm{I}} & 0 & 0 \end{bmatrix}\right]_P$$

$$((\overset{*}{U}{}_\Omega{}^{ab}(P))) \equiv ((\overset{*}{U}{}_\Omega{}^{[ab]}(P))) =$$

$$\left[\begin{bmatrix} 0 & 0 & E_1{}^{\mathrm{II}} & E_1{}^{\mathrm{I}} & E_1{}^{\mathrm{III}} & E_1{}^{\mathrm{IV}} \\ 0 & 0 & E_2{}^{\mathrm{II}} & E_2{}^{\mathrm{I}} & E_2{}^{\mathrm{III}} & E_2{}^{\mathrm{IV}} \\ 0 & 0 & E_3{}^{\mathrm{II}} & E_3{}^{\mathrm{I}} & E_3{}^{\mathrm{III}} & E_3{}^{\mathrm{IV}} \\ 0 & 0 & E_4{}^{\mathrm{II}} & E_4{}^{\mathrm{I}} & E_4{}^{\mathrm{III}} & E_4{}^{\mathrm{IV}} \\ \epsilon(-1) & 0 & 0 & 0 & 0 & 0 \\ 0 & \epsilon(1) & 0 & 0 & 0 & 0 \end{bmatrix}\right]_P$$

$(\epsilon = \pm)$

and the scalars associated with a vector $v^\lambda(P)$

$$\overset{\cdot}{v}{}^5 = \overset{\cdot}{v}{}^6 = \sqrt{h_{\lambda\mu} v^\lambda v^\mu} \qquad \text{at } P$$

and put

$$(a) \quad v_b{}^a(P) \equiv E_{\lambda b}{}^a(P) v^\lambda(P)$$

$$(b) \quad \overset{*}{v}{}^{ab}(P) \equiv \overset{*}{U}{}_\Omega{}^{ab}(P) v^\Omega(P) \tag{4}$$

The numbers $v_b{}^a(P)$ may be looked upon as a set of homogeneous components of a tensor in a projective three-dimensional space which we denote by $S_3(P)$. Assume first that $v^\lambda(P)$ is anisotropic. Then the tensor $v_b{}^a(P)$ defines a biaxial involution in $S_3(P)$, whose (mutually skew) axes have the Plücker point coordinates $\overset{*}{v}{}^{ab}(P)$ given by (4*b*). If however, $v^\lambda(P)$ is isotropic then the lines (4*b*) coincide $\overset{+}{v}{}^{ab} = \overset{-}{v}{}^{ab} \equiv v^{ab}$ and constitute a linear

hyperbolic line congruence $C(P)$ in $S_3(P)$ as $v^\lambda(P)$ moves. The axes of $C(P)$ have the Plücker point coordinates

$$^1A^{ab}\ (0\ 1\ 0\ 0\ 0\ 0), \qquad ^2A^{ab}\ (1\ 0\ 0\ 0\ 0\ 0).$$

The locus of all lines $^\lambda v^{ab}(P)$ and all lines of $C(P)$ is a linear line complex $K(P)$ in $S_3(P)$. Its coordinates are

$$K^{ab} \equiv {}^1A^{ab} + {}^2A^{ab}.$$

The space $S_3(P)$ will be referred to as the *local spinor space* attached to $T_4(P)$. Every point $S^a(P)$ of $S_3(P)$ will be referred to as a spinor in $S_3(P)$. The change of frame of reference in $S_3(P)$ is

$$'S^a(P) = \Lambda_b{}^a(P)\ S^b(P) \tag{5}$$

where the Λ's are determined by (2)

$$((\Lambda_b{}^a(P))) \equiv \left[\left[\begin{matrix} \alpha e^{\psi+i\varphi} & 0 & 0 & 0 \\ 0 & \alpha e^{-(\psi+i\varphi)} & 0 & 0 \\ 0 & 0 & \frac{1}{\alpha} e^{\psi-i\varphi} & 0 \\ 0 & 0 & 0 & \frac{1}{\alpha} e^{-(\psi-i\varphi)} \end{matrix}\right]\right]_P \tag{6}$$

and $\alpha = 1, i$. We denote by $\overset{-1}{\Lambda}_b{}^a(P)$ the elements of the matrix inverse to (6). The line $v^{ab}(P)$ in $S_3(P)$ corresponding to an isotropic point in $I_3(P)$ has the equations

$$\begin{aligned} S^3(P)v^{\mathrm{I}}(P) + S^4(P)v^{\mathrm{III}}(P) &= 0 \\ S^1(P)v^{\mathrm{II}}(P) + S^2(P)v^{\mathrm{III}}(P) &= 0 \end{aligned} \tag{7}$$

where $S^a(P)$ is a generic *not focal* point of $v^{ab}(P)$. The components $v^{\mathrm{I}}(P)$ and $v^{\mathrm{II}}(P)$ being complex conjugate (while $v^{\mathrm{III}}(P)$ and $v^{\mathrm{IV}}(P)$ are real) we have from (7)

$$(a) \qquad \frac{\overline{S^1(P)}}{\overline{S^2(P)}} = \frac{S^3(P)}{S^4(P)} \tag{8}$$

or sometimes even

$$(b) \qquad \overline{S^1(P)} = S^3(P) \qquad \overline{S^2(P)} = S^4(P). \tag{8}$$

A spinor $S^a(P)$ satisfying (8*a*) but not (8*b*) will be referred to as an analytical spinor. A spinor satisfying (8*b*) will be termed a physical spinor. This distinction is preserved by (6).

4. *The Spinor Space* $\sum_3$.—Every point P of X_4 gives rise to a local

spinor space $S_3(P)$, with the congruence $C(P)$ and the complex $K(P)$. The axes ${}^1A$ and ${}^2A$ have the same Plücker coordinates in every $S_3(P)$. Therefore we may identify them with two arbitrary skew lines of a three-dimensional projective space $\sum_3$. In doing so we superpose all spaces $S_3(P)$ on $\sum_3$ in a manner just mentioned. The space $\sum_3$ will be referred to as the *spinor space attached to* X_4. As the point P moves along X_4 the map of an isotropic field $v^A = E_\lambda{}^A v^\lambda$ is the set of lines

$$v^{ab}(0\ 0\ v^{II}\ v^{I}\ v^{III}\ v^{IV}) \tag{9}$$

in $\sum_3$. Hence these lines belong to the congruence with the axes ${}^1A$, ${}^2A$. In other words, the congruence $C(P)$ of $S_3(P)$ is identical with the just mentioned congruence C in $\sum_3$.[5] The same is true for the complex $K(P)$ of $S_3(P)$. It is identical with the complex ${}^1A^{ab} + {}^2A^{ab}$ in $\sum_3$, $K(P) \equiv K$.

5. *Connections of* $\sum_3$.—The superposition of all $S_3(P)$ on $\sum_3$ results in a set of connections in $\sum_3$, which we are going to investigate: Let P and Q be two points of X_4, $S_3(P)$ and $S_3(Q)$ their corresponding local spinor spaces (both superposed on $\sum_3$). Consider a projectivity $P_a{}^b(P|Q)$ which maps a spinor $S^a(P)$ of $S_3(P)$ on a spinor

$$S^a(P|Q) = P_b{}^a(P|Q)S^b(P) \tag{10}$$

of $S_3(Q)$. We assume that this projectivity does not depend on the reference frame, is not degenerate and reduces to identity for $P = Q$. Moreover, we assume that a physical spinor is mapped on a physical spinor by (11). Using the fact that $C = C(P)$ for all P of X_4 *we require also that the axes* ${}^1A$ *and* ${}^2A$ *of* C *be reproduced by* (11). Keeping Q fixed and putting

$$\Pi_{b\lambda}{}^a = \left(\frac{\partial}{\partial x^\lambda} P_b{}^a(P|Q)\right)_{P=Q}$$

we obtain in $\Pi_{b\lambda}{}^a$ a connection

$$'\Pi_{b\lambda}{}^a = \Lambda_c{}^a \overset{-1}{\Lambda}{}_b{}^d \Pi_{d\lambda}{}^c + \Lambda_c{}^a \frac{\partial}{\partial x^\lambda} \overset{-1}{\Lambda}{}_b{}^c \tag{11}$$

so that in particular

$$(a) \qquad \Gamma_{b\lambda}{}^a \equiv \Pi_{b\lambda}{}^a - \frac{1}{4}\delta_b{}^a \Pi_{c\lambda}{}^c \tag{12}$$

is also a connection and $\Gamma_{a\lambda}{}^a = 0$. This connection is given by the matrix

$$(b) \qquad ((\Gamma_{b\lambda}{}^a)) = \begin{bmatrix}\begin{bmatrix} a_\lambda & b_\lambda & 0 & 0 \\ c_\lambda & -a_\lambda & 0 & 0 \\ 0 & 0 & \bar{a}_\lambda & \bar{b}_\lambda \\ 0 & 0 & \bar{c}_\lambda & -\bar{a}_\lambda \end{bmatrix}\end{bmatrix} \tag{12}$$

where a_λ, b_λ, c_λ are arbitrary vector functions of P and $\bar{a}_\lambda$, $\bar{b}_\lambda$, $\bar{c}_\lambda$ their complex conjugate.

6. The Connection of $\sum_3$.—Let us denote by Ξ_μ the symbol of the covariant derivative with respect to (12) and assume that it operates in X_4 by means of the Christoffel symbols of $h_{\lambda\mu}$. In order to define the vectors a_λ, b_λ, c_λ in a proper way we require that the relationship (4*a*) be not discarded by the operator $dx^\mu \Xi_\mu$ (dx arbitrary).

$$(a) \qquad \Xi_\mu E_{\lambda a}{}^{b} = 0. \tag{13}$$

Using now the definition of the matrix $E_{\lambda_a}{}^{b}$ as well as (12*b*) we obtain from (13*a*)

$$\begin{aligned}(b) \qquad & -E_{\mathrm{III}}{}^{\lambda}\,\Xi_\mu\,E_\lambda{}^{\mathrm{III}} + E_{\mathrm{II}}{}^{\lambda}\,\Xi_\mu\,E_\lambda{}^{\mathrm{II}} = 2a_\mu \\ & -E_{\mathrm{IV}}{}^{\lambda}\,\Xi_\mu\,E_\lambda{}^{\mathrm{I}} = E_{\mathrm{II}}{}^{\lambda}\,\Xi_\mu\,E_\lambda{}^{\mathrm{III}} = b_\mu \\ & E_{\mathrm{III}}{}^{\lambda}\,\Xi_\mu\,E_\lambda{}^{\mathrm{II}} = -E_{\mathrm{I}}{}^{\lambda}\,\Xi_\mu\,E_\lambda{}^{\mathrm{IV}} = c_\mu.\end{aligned} \tag{13}$$

Interchanging the indices I and II in (13*b*) we obtain $2\bar{a}_\mu$, $\bar{b}_\mu$, $\bar{c}_\mu$. Substituting these values into (12*b*) we obtain a *uniquely defined connection* $\Gamma_{b\lambda}{}^{a}$ *of* $\sum_3$, *for which* (13*a*) *holds*. It is to be remarked that

$$\Xi_\mu\,{}^{1}A^{ab} = \Xi_\mu\,{}^{2}A^{ab} = \Xi_\mu\,K^{ab} = 0$$

as it was to be expected.

* Prepared under Army Contract DA 33-008 ORD-467.

[1] Einstein, A., *The Meaning of Relativity*, Appendix II, 1950.

[2] The Greek indices run from 1 to 4, the small Latin indices from 1 to 4, the capital latin indices from I to IV. Moreover, the index Ω in section 3 runs from 1 to 6.

[3] The proofs of the statements of sections 2, 3, 5 are based on the results of the following papers:

Section 2: Hlavatý, V., "The Elementary Basic Principles of the Unified Theory of Relativity. A," *J. Rat. Mech. Anal.* **1**, 539–562 (1952). The statement about the structure of $g_{\lambda\mu}$ is taken from this paper.

Section 3: Hlavatý, V., "Spinor Space and Line Geometry," *Can. J. Math.*, **3**, No. 4, 442–459 (1951) and "Spinor Space and Line Geometry II," *J. Rat. Mech. Anal.* **1**, 321–339 (1952).

Section 5: Bortolotti, E., and Hlavatý, V., Contributi alla teoria delle connessioni," *Annali di Matematica*, **15**, Series 4, (1–71) (1936–37).

A spinor connection based on projective theory of relativity is dealt with in Veblen, O., and Taub, A. H., "Projective Differentiation of Spinors," PROC. NATL. ACAD. SCI. **20**, 85–92 (1934).

[4] The superscript b (the subscript a) in the first matrix denotes the rows (columns). In the second matrix the columns are labeled by the superscripts ab which take the values

12 34 23 14 31 24

in this order.

[5] The lines (9) in Σ_3 belonging to an isotropic *field* v^λ are the rulings of the congruence C. The same congruence regarded as $C(P)$ in $S_3(P)$ is constituted by the rulings which map on $S_3(P)$ all isotropic points $v^\nu(P)$ for P fixed.

CONNECTIONS BETWEEN EINSTEIN'S TWO UNIFIED THEORIES OF RELATIVITY

BY VÁCLAV HLAVATÝ

GRADUATE INSTITUTE FOR APPLIED MATHEMATICS, INDIANA UNIVERSITY, BLOOMINGTON, INDIANA*

Communicated by T. Y. Thomas, March 26, 1953

1. The Problem to Be Solved.—In his three earlier papers[1] Einstein has based his unified theory on a complex tensor

$$(a) \qquad {}^*g_{\lambda\mu} = h_{\lambda\mu} + ik_{\lambda\mu}$$
$$h_{\lambda\mu} = h_{(\lambda\mu)}, \qquad k_{\lambda\mu} = k_{[\lambda\mu]} \tag{1}$$

and on the complex connection ${}^*\Gamma^{\nu}_{\lambda\mu}$ which satisfies the field equations

$$(a) \qquad \partial_\omega {}^*g_{\lambda\mu} = {}^*\Gamma^{\alpha}_{\lambda\omega}\,{}^*g_{\alpha\mu} + {}^*\Gamma^{\alpha}_{\omega\mu}\,{}^*g_{\lambda\alpha}. \tag{2}$$

The tensor ${}^*g_{\lambda\mu}$ being obviously hermitian symmetric

$$(a) \qquad \overline{{}^*g_{\mu\lambda}} = {}^*g_{\lambda\mu} \tag{3}$$

one assumes also the hermitian symmetry of the connection

$$(b) \qquad {}^*\bar{\Gamma}^{\nu}_{\mu\lambda} = {}^*\Gamma^{\nu}_{\lambda\mu}. \tag{3}$$

In his recent book[2] Einstein based his unified theory on a real tensor

$$(b) \qquad g_{\lambda\mu} = h_{\lambda\mu} + k_{\lambda\mu}$$
$$h_{\lambda\mu} = h_{(\lambda\mu)}, \qquad k_{\lambda\mu} = k_{[\lambda\mu]} \tag{1}$$

and on a real connection $\Gamma^{\nu}_{\lambda\mu}$ which satisfies the field equations

$$(b) \qquad \partial_\omega g_{\lambda\mu} = \Gamma^{\alpha}_{\lambda\omega} g_{\alpha\mu} + \Gamma^{\alpha}_{\omega\mu} g_{\lambda\alpha}. \tag{2}$$

Let us identify the tensors $h_{\lambda\mu}$ in (1a, b) as well as the tensors $k_{\lambda\mu}$ in (1a, b). The main scope of this paper is to find the relationship between ${}^*\Gamma^{\nu}_{\lambda\mu}$ given by (2a) and $\Gamma^{\nu}_{\lambda\mu}$ given by (2b).

2. First Step.—Let us put

$${}^*\Gamma^{\nu}_{\lambda\mu} = {}^*\Lambda^{\nu}_{\mu\lambda} + i\,{}^*S_{\lambda\mu}{}^{\nu} \tag{4}$$

where ${}^*\Lambda$ and ${}^*S$ are real. The condition (3*b*) yields

$${}^*\Lambda^{\nu}_{\lambda\mu} = {}^*\Lambda^{\nu}_{(\mu\lambda)}, \quad {}^*S_{\lambda\mu}{}^{\nu} = {}^*S_{[\lambda\mu]}{}^{\nu}. \tag{5}$$

The underlying space being real and referred to real coordinates we see at once from the transformation rule of the ${}^*\Gamma$'s as well as from (5) that ${}^*\Lambda^{\nu}_{\lambda\mu}$ is a (symmetric) connection while ${}^*S_{\lambda\mu}{}^{\nu}$ is a skew-symmetric tensor and in particular

$$(a) \qquad {}^*\Gamma_{[\lambda\mu]}{}^{\nu} = i\,{}^*S_{\lambda\mu}{}^{\nu}. \tag{6}$$

In the next section we will compute ${}^*\Lambda_{\lambda\mu}{}^{\nu}$ as well as ${}^*S_{\lambda\mu}{}^{\nu}$.

3. Covariant Derivative ${}^*\mathring{D}_{\mu}$. If we denote by ${}^*\mathring{D}_{\mu}$ the symbol of the covariant derivative with respect to the symmetric connection ${}^*\Lambda_{\lambda\mu}^{\nu}$, then (2a) splits in two real sets namely

$$\begin{aligned} (a) \qquad & {}^*\mathring{D}_{\omega} h_{\lambda\mu} = -2\,{}^*S_{\omega(\lambda}{}^{\alpha} k_{\mu)\alpha} \\ (b) \qquad & {}^*\mathring{D}_{\omega} k_{\lambda\mu} = -2\,{}^*S_{\omega[\lambda}{}^{\alpha} h_{\mu]\alpha}. \end{aligned} \tag{7}$$

Introducing the symbol ∇_{μ} of the covariant derivative with respect to the Christoffel symbols $\left\{ {}^{\nu}_{\lambda\mu} \right\}$ of $h_{\lambda\mu}$[3] and putting

$${}^*\Lambda_{\lambda\mu}^{\nu} = \left\{ {}^{\nu}_{\lambda\mu} \right\} + {}^*U^{\nu}{}_{\lambda\mu} \tag{8}$$

we may rewrite the equations (7) as follows

$$\begin{aligned} (a) \qquad & {}^*U^{\alpha}{}_{\lambda\omega}\, h_{\alpha\mu} + {}^*U^{\alpha}{}_{\mu\omega}\, h_{\lambda\alpha} = 2\,{}^*S_{\omega(\lambda}{}^{\alpha} k_{\mu)\alpha} \\ (b) \qquad & \nabla_{\omega} k_{\lambda\mu} = {}^*U^{\alpha}{}_{\lambda\omega}\, k_{\alpha\mu} + {}^*U^{\alpha}{}_{\mu\omega}\, k_{\lambda\alpha} + 2\,{}^*S_{\omega[\mu}{}^{\alpha} h_{\lambda]\alpha}\,. \end{aligned} \tag{9}$$

Let us denote by $(\mu\lambda\omega)$ the equation (9a). If we compute

$$(\mu\lambda\omega) + (\lambda\omega\mu) - (\omega\mu\lambda)$$

we obtain

$$(a) \qquad {}^*U^{\nu}{}_{\lambda\mu} = -2\, h^{\nu\alpha}\, {}^*S_{\alpha(\lambda}{}^{\beta} k_{\mu)\beta}. \tag{10}$$

On the other hand if we introduce the tensor

$$K_{\omega\mu\nu} = \nabla_{\omega} k_{\nu\mu} + \nabla_{\mu} k_{\omega\nu} + \nabla_{\nu} k_{\omega\mu}$$

we have from (10) and (9b)[4]

$$2\,{}^*S_{\omega\mu\nu} = K_{\omega\mu\nu} - 4k_{[\omega}{}^{\beta}\,{}^*S_{\mu]\beta\gamma} k_{\nu}{}^{\gamma} - 4k_{[\omega}{}^{[\beta} k_{\mu]}{}^{\gamma]}\,{}^*S_{\nu\beta\gamma}. \tag{11}$$

If we put

$${}^*X^{\xi\eta\zeta}_{\omega\mu\nu} \equiv \delta^{[\xi}_{[\omega}\delta^{\eta]}_{\mu]}\,\delta^{\zeta}_{\nu} + 2k_{[\omega}{}^{[\eta}\delta^{\xi]}_{\mu]}\, k_{\nu}{}^{\zeta} + 2\delta^{[\xi}_{\nu}\, k_{[\omega}{}^{\eta]}\, k_{\mu]}{}^{\zeta}$$

then we may write (11a) in a more condensed form

$$(a) \qquad {}^*S_{\xi\eta\zeta}\,{}^*X^{\xi\eta\zeta}_{\omega\mu\nu} = \frac{1}{2}\, K_{\omega\mu\nu}. \tag{12}$$

4. The System (2b).—The system (2b) has been solved for the Γ's in a previous paper.[5] In particular we have obtained the solution in the form

$$\Gamma_{\lambda\mu}^{\nu} = \left\{ {}^{\nu}_{\lambda\mu} \right\} + U^{\nu}{}_{\lambda\mu} + S_{\lambda\mu}{}^{\nu} \tag{13}$$

where

$$(b) \qquad U^{\nu}{}_{\lambda\mu} = 2h^{\nu\alpha} S_{\alpha(\lambda}{}^{\beta} k_{\mu)\beta} \tag{10}$$

$$(b) \qquad S_{\lambda\mu}{}^{\nu} = \Gamma^{\nu}_{[\lambda\mu]}. \tag{6}$$

The tensor $S_{\lambda\mu\nu}$ satisfies the equations

$$(b) \qquad S_{\xi\eta\zeta} X^{\xi\eta\zeta}_{\omega\mu\nu} = \frac{1}{2} K_{\omega\mu\nu} \tag{12}$$

with

$$X^{\xi\eta\zeta}_{\omega\mu\nu} = \delta^{[\xi}_{[\omega} \delta^{\eta]}_{\mu]} \delta^{\zeta}_{\nu} - 2k_{[\omega}{}^{[\eta} \delta^{\xi]}_{\mu]} k_{\nu}{}^{\zeta} - 2\delta^{[\xi}_{\nu} k_{[\omega}{}^{\eta]} k_{\mu]}{}^{\zeta} .$$

Denote by g, h, k the determinants of $g_{\lambda\mu}$, $h_{\lambda\mu}$ and $k_{\lambda\mu}$, respectively. (We already assumed $h \neq 0$.) The system (12b) admits one solution $S_{\xi\eta\zeta}$ if and only if $\frac{g}{h}\frac{k}{h} \neq 0$ and whenever $k = 0$, if and only if $\frac{g}{h}\left(\frac{g}{h} - 2\right) \neq 0$.[6] As far as the system (12a) is concerned it is not difficult to find necessary and sufficient conditions in terms of the $^*X$'s that it admit one solution only.[7] We assume that (12a) admits only one solution.

5. *Relationship Between* $\Gamma^{\nu}_{\lambda\mu}$ *and* $^*\Gamma^{\nu}_{\lambda\mu}$.—Put

$$A_{\omega\mu\nu} \equiv 2\, S_{\omega\mu\nu} + 2\, {}^*S_{\omega\mu\nu}$$
$$B_{\omega\mu\nu} \equiv 2\, S_{\omega\mu\nu} - 2\, {}^*S_{\omega\mu\nu} .$$

Then we have easily from (12a, b)

$$\begin{aligned} A_{\omega\mu\nu} &= 2K_{\omega\mu\nu} + 2k_{[\omega}{}^{\beta} B_{\mu]\beta\gamma} k_{\nu}{}^{\gamma} + 2k_{[\omega}{}^{[\beta} k_{\mu]}{}^{\gamma]} B_{\nu\beta\gamma} \\ B_{\omega\mu\nu} &= \qquad\qquad 2k_{[\omega}{}^{\beta} A_{\mu]\beta\gamma} k_{\nu}{}^{\gamma} + 2k_{[\omega}{}^{[\beta} k_{\mu]}{}^{\gamma} A_{\nu\beta\gamma} . \end{aligned} \tag{14}$$

Substituting from the second of (14) into the first one we obtain

$$A_{\xi\eta\zeta} Z^{\xi\eta\zeta}_{\omega\mu\nu} = 2K_{\omega\mu\nu} \tag{15}$$

where

$$Z^{\xi\eta\zeta}_{\omega\mu\nu} = \delta^{[\xi}_{[\omega} \delta^{\eta]}_{\mu]} \delta^{\zeta}_{\nu} + {}'Z^{\xi\eta\zeta}_{\omega\mu\nu}$$

and ${}'Z^{\xi\eta\zeta}_{\omega\mu\nu} = {}'Z^{[\xi\eta]\zeta}_{[\omega\mu]\nu}$ is a well-defined sum of products of four components $k_{\mu\lambda}$. If the system (12a) admits only one solution $^*S_{\omega\mu\nu}$ and moreover if the system (12b) admits only one solution $S_{\omega\mu\nu}$ then (15) admits only one solution so that

$$^*S_{\alpha\beta\gamma} = -S_{\alpha\beta\gamma} + T^{\omega\mu\nu}_{\alpha\beta\gamma} K_{\omega\mu\nu}. \tag{16}$$

Here the tensor $T^{\omega\mu\nu}_{\alpha\beta\gamma} = T^{[\omega\mu]\nu}_{[\alpha\beta]\gamma}$ is uniquely defined by

$$Z^{\xi\eta\zeta}_{\omega\mu\nu} T^{\omega\mu\nu}_{\alpha\beta\gamma} = \delta^{[\omega}_{[\alpha} \delta^{\mu]}_{\beta]} \delta^{\nu}_{\gamma}.$$

Hence if we consider (6) and (8), i.e.,

$$^*\Gamma^{\nu}_{\lambda\mu} = \left\{ {\nu \atop \lambda\mu} \right\} + {}^*U^{\nu}{}_{\lambda\mu} + i\, {}^*S_{\lambda\mu}{}^{\nu}$$

and substitute into it from (16) we obtain by virtue of (10) and (12b)

$$\begin{aligned} {}^*\Gamma^{\nu}_{\lambda\mu} &= \Gamma^{\nu}_{\lambda\mu} - S_{\lambda\mu}{}^{\nu} + 2h^{\alpha\nu} K_{\xi\eta\zeta} k_{(\mu}{}^{\beta} T^{\xi\ \eta\zeta}_{\lambda)\alpha\beta} - i(S_{\lambda\mu}{}^{\nu} - h^{\alpha\nu} K_{\xi\eta\zeta} T^{\xi\ \eta\zeta}_{\lambda\mu\alpha}) \\ &= \Gamma^{\nu}_{\lambda\mu} - (i+1) S_{\lambda\mu}{}^{\nu} + h^{\alpha\nu} K_{\xi\eta\zeta} [2k_{(\mu}{}^{\beta} T^{\xi\ \eta\zeta}_{\lambda)\alpha\beta} + i T^{\xi\eta\zeta}_{\lambda\mu\alpha}]. \end{aligned}$$

This formula establishes the relationship between the connections ${}^*\Gamma^{\nu}_{\lambda\mu}$ and $\Gamma^{\nu}_{\lambda\mu}$.

* Prepared under Army Contract DA 33-008 ORD-467.

[1] Einstein, Albert, "A Generalization of the Relativistic Theory of Gravitation," *Ann. Math.* (2), **46,** 578–584 (1945) and "A Generalized Theory of Gravitation," *Rev. Modern Physics,* **20,** 35–39 (1948). Einstein, A., and Straus, E. G., "A Generalization of the Relativistic Theory of Gravitation, II," *Ann. Math.* (2), **47,** 731–741 (1946).

[2] Einstein, A., *The Meaning of Relativity,* Princeton University Press, 1950, Appendix II.

[3] The existence of $\left\{ {\nu \atop \lambda\mu} \right\}$ is possible if and only if the determinant h of $h_{\lambda\mu}$ is different from zero.

[4] The indices are raised and lowered by means of $h^{\lambda\nu}$ and $h_{\lambda\mu}$.

[5] Hlavatý, V., "The Elementary Basic Principles of the Unified Theory of Relativity. B," *J. Rat. Mech. Anal.* (1), **2,** 1–52 (1953).

[6] Hlavatý, V., and Säenz, A. W., "Uniqueness Theorems in the Unified Theory of Relativity," to be published in the *J. Rat. Mech. Anal.*

[7] These conditions may be formulated in a similar way as it has been done for the system (12*b*) in the paper mentioned in Ref. 5.

ON THE CONTRACTION OF GROUPS AND THEIR REPRESENTATIONS

BY E. INONU AND E. P. WIGNER

PALMER PHYSICAL LABORATORY, PRINCETON UNIVERSITY

Communicated April 21, 1953

Introduction.—Classical mechanics is a limiting case of relativistic mechanics. Hence the group of the former, the Galilei group, must be in some sense a limiting case of the relativistic mechanics' group, the representations of the former must be limiting cases of the latter's representations. There are other examples for similar relations between groups. Thus, the inhomogeneous Lorentz group must be, in the same sense, a limiting case of the de Sitter groups. The purpose of the present note is to investigate, in some generality, in which sense groups can be limiting cases of other groups (Section I), and how their representations can be obtained from the representations of the groups of which they appear as limits (Section II). Section III deals briefly with the transition from inhomogeneous Lorentz group to Galilei group. It shows in which way the representation up to a factor of the Galilei group, embodied in

the Schrödinger equation, appears as a limit of a representation of the inhomogeneous Lorentz group and also gives the reason why no physical interpretation is possible for the real representations of that group.

I. Contraction of Groups

Let us consider an arbitrary Lie group with n parameters a^i and infinitesimal operators I_i. These shall be given, as usual by

$$I_i = \lim_{h \to 0} \frac{g(he_i) - g(0)}{h} \tag{1}$$

where 0 are the parameters of the unit element and e_i differs from 0 by a unit increase of a_i. The I_j are skew hermitean matrices if the group consists of unitary matrices; instead of them one often uses the hermitean quantities iI_j. However, all our equations remain somewhat simpler if expressed in terms of the I_j. The structure constants C are defined by

$$[I_i, I_j] = \sum_k C_{ij}^k I_k \tag{2}$$

If we subject the I_i to a linear homogeneous non-singular transformation, the C will be replaced by other constants. These are obtained from the C by contragradient transformations of its upper and lower indices. However, such a transformation has, naturally, no effect on the structure of the group. Let us denote the transformation in question by

$$J_\nu = \sum_i I_i U^i{}_\nu. \tag{3}$$

It corresponds to the transformation

$$a^i = \sum_\kappa U^i{}_\kappa b^\kappa \tag{3a}$$

according to which the J are obtained by the same equation (1) as the I, except that the e_i have to be replaced in it by a similar quantity, defined with respect to the b.

The above transformation may lead to a new group only if the matrix U of (3) is singular. We shall call the operation of obtaining a new group by a singular transformation of the infinitesimal elements of the old group a *contraction* of the latter. The reason for this term will become clear below. The singular matrix will be a limiting case of a non-singular matrix. The latter will depend linearly on a parameter ϵ which will tend to zero:

$$U^i{}_\nu = u^i{}_\nu + \epsilon w^i{}_\nu. \tag{4}$$

For $0 < \epsilon < \epsilon_0$ the determinant of (4) is different from zero, it vanishes for $\epsilon = 0$.

We shall transform (4) into a normal form by a non-singular and ϵ

independent transformation of both the J_ν and I_i. If the matrices of the corresponding transformations are denoted by α and β, respectively, u and w will be replaced by $\beta u \alpha^{-1}$ and $\beta w \alpha^{-1}$. It is possible, by such a transformation, to give u and w the form

$$u = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix} \qquad w = \begin{Vmatrix} v & 0 \\ 0 & 1 \end{Vmatrix}. \tag{5}$$

The number of rows and columns in the unit matrix in u, and in v, is equal to the rank r of u. It is advantageous to label the transformed J and I with a pair of indices, the first referring to the subdivision of u given in (5), the second specifying the various J and I within that subdivision. Hence, (3) assumes the form

$$\begin{aligned} J_{1\nu} &= I_{1\nu} + \epsilon \sum_{\mu=1}^{r} v_{\mu\nu} I_{1\mu} \qquad (\nu = 1, 2, \ldots, r) \\ J_{2\nu} &= \epsilon I_{2\nu}. \qquad\qquad\qquad (\nu = 1, 2, \ldots, n - r) \end{aligned} \tag{6}$$

The corresponding transformation of the group parameters is

$$\begin{aligned} a_{1\nu} &= b_{1\nu} + \epsilon \sum_{\mu=1}^{r} v_{\nu\mu} b_{1\mu} \qquad (\nu = 1, 2, \ldots, r) \\ a_{2\nu} &= \epsilon b_{2\nu}. \qquad\qquad\qquad (\nu = 1, 2, \ldots, n - r) \end{aligned} \tag{6a}$$

It is well to remember that the parameters a lead to the infinitesimal elements I, the parameters b to the J. The last equation shows that a given set of parameters b correspond, with decreasing ϵ, to smaller and smaller values of the parameters a_2. In the limit $\epsilon = 0$ (if such a limit exists), one will have contracted the whole group to an infinitesimally small neighborhood of the group defined by the $a_{1\nu}$ alone. This justifies the name given to the process considered.

The transformation of the infinitesimal elements which we carried out also changes the structure constants and we shall write for (2)

$$[I_{\alpha\nu}, I_{\beta\mu}] = \sum_{\kappa=1}^{r} C_{\alpha\nu,\,\beta\mu}{}^{1\kappa} I_{1\kappa} + \sum_{\kappa=1}^{n-r} C_{\alpha\nu,\,\beta\mu}{}^{2\kappa} I_{2\kappa} \tag{7}$$

wherein α and β can assume the values 1 and 2. This gives for

$$\begin{aligned} [J_{1\nu}, J_{1\mu}] &= [I_{1\nu}, I_{1\mu}] + \epsilon \sum (v_{\nu\nu'}\delta_{\mu\mu'} + \delta_{\nu\nu'}v_{\mu\mu'} + \epsilon v_{\nu\nu'}v_{\mu\mu'})[I_{1\nu'}, I_{1\mu'}] \\ &= \sum_{\kappa} C_{1\nu,\,1\mu}{}^{1\kappa} J_{1\kappa} + \frac{1}{\epsilon} \sum_{\kappa} C_{1\nu,\,1\mu}{}^{2\kappa} J_{2\kappa} + 0(1). \end{aligned}$$

Hence, if the commutator of $J_{1\nu}$ and $J_{1\mu}$ is to converge, as $\epsilon \to 0$, to a linear combination of the J, the structure constants

$$C_{1\nu,\,1\mu}{}^{2\kappa} = 0 \tag{8}$$

i.e., the $I_{1\nu}$ must span a subgroup. On the other hand, if this happens to be the case, the structure constants will converge to definite values $c_{\alpha\nu,\,\beta\mu}{}^{\gamma\kappa}$ as $\epsilon \to 0$

$$\begin{aligned} c_{1\nu,\,1\mu}{}^{1\kappa} &= C_{1\nu,\,1\mu}{}^{1\kappa} & c_{1\nu,\,1\mu}{}^{2\kappa} &= C_{1\nu,\,1\mu}{}^{2\kappa} = 0 \\ c_{1\nu,\,2\mu}{}^{1\kappa} &= 0 & c_{1\nu,\,2\mu}{}^{2\kappa} &= C_{1\nu,\,2\mu}{}^{2\kappa} \\ & c_{2\nu,\,2\mu}{}^{1\kappa} = c_{2\nu,\,2\mu}{}^{2\kappa} = 0. \end{aligned} \tag{9}$$

These structure constants satisfy Jacobi's identities since the structure constants for the J do this for non-vanishing ϵ. We shall say that the above operation is a contraction of the group with respect to the infinitesimal elements $I_{1\nu}$ or that the infinitesimal elements $I_{2\mu}$ are contracted. We then have, from (9).

THEOREM 1. *Every Lie group can be contracted with respect to any of its continuous subgroups and only with respect to these.* The subgroup with respect to which the contraction is undertaken will be called S. *The contracted infinitesimal elements form an abelian invariant subgroup of the contracted group. The subgroup S with respect to which the contraction was undertaken is isomorphic with the factor group of this invariant subgroup. Conversely, the existence of an abelian invariant subgroup and the possibility to choose from each of its cosets an element so that these form a subgroup S, is a necessary condition for the possibility to obtain the group from another group by contraction.*

It is easy to visualize now the effect of the contraction on the whole group. The subgroup S with respect to which the contraction is undertaken remains unchanged and it is advantageous to choose the group parameters in such a way that $a_{2\nu} = 0$ throughout S. Then (6a) can be replaced by

$$a_{1\nu} = b_{1\nu} \qquad a_{2\nu} = \epsilon b_{2\nu} \tag{6b}$$

and this can be assumed to be valid throughout the whole group, not only in the neighborhood of the unit element. As ϵ decreases, a fixed range of the parameter b will describe an increasingly small surrounding of S. As ϵ tends to 0, the range of the $b_{2\nu}$ will become infinite and describe only those group elements which differ infinitesimally from the elements of S. The elements which are in the neighborhood of the unit element of the original group but have finite parameters $b_{2\nu}$ will commute and form the aforementioned commutative invariant subgroup. Naturally, the elements of this invariant subgroup will not commute, in general, with the elements of the subgroup S: the change of the parameters $a_{2\nu} = \epsilon b_{2\nu}$ will be, upon transformation by finite elements of S, of the same order of magnitude as these parameters themselves. Naturally, the convergence

of the original group toward the contracted group is a typically non-uniform convergence.

Every Lie group can be contracted with respect to any of its one parametric subgroups. If the three dimensional rotation group is contracted in this way, one obtains the Euclidean group for two dimensions. Contraction of the homogeneous Lorentz group with respect to the subgroup which leaves the time coordinate invariant yields the homogeneous Galilei group, contraction of the inhomogeneous Lorentz group with respect to the group generated by spatial rotations and time displacements yields the full Galilei group. Contraction of the de Sitter groups yields the inhomogeneous Lorentz group. It should be remarked, finally, that if a group, obtained by contraction of another group with respect to the subgroup S, is contracted again with respect to S, the second contraction remains without effect.

The above considerations show a certain similarity with those of I. E. Segal.[1] However, Segal's considerations are more general than ours as he considers a sequence of Lie groups the structure constants of which converge toward the structure constants of a non-isomorphic group. In the above, we have considered only one Lie group but have introduced a sequence of coordinate systems therein and investigated the limiting case of these coordinate systems becoming singular. As a result of our problem being more restricted, we could arrive at more specific results.

II. Contraction of Representations

If one applies the transformation (6) to the infinitesimal elements of a representation of the group to be contracted, and lets ϵ tend to zero, the $J_{2\nu}$ will also tend to zero. The representation will become isomorphic to the representation of the subgroup S, i.e., will be a representation of the factor group of the invariant subgroup. In order to obtain a faithful representation, one must *either subject the $J_{2\nu}$ to an ϵ dependent transformation, or consider the $J_{2\nu}$ which correspond to different representations*, e.g., go to higher and higher dimensional representations as ϵ decreases. We shall give examples for both procedures.

(*a*) *Representations of the Contracted Group by Means of ϵ Dependent Transformations.*—The first procedure is applicable only if the infinitesimal elements are not bounded operators, i.e., as far as irreducible representations are concerned, only if the group is not compact. The simplest non-compact non-commutative group is that of the linear transformations $x' = e^{a}x + \beta$. The general group element is $O_{\alpha,\beta} = T(\beta)R_{\alpha}$ with the group relations $T(\beta)T(\beta') = T(\beta + \beta')$; $R_{\alpha}R_{\alpha'} = R_{\alpha+\alpha'}$ and $R_{\alpha}T(\beta) = T(e^{a}\beta)R_{\alpha}$. The only faithful irreducible unitary representation of this group can be given in the Hilbert space of square integrable functions of $0 < x < \infty$

$$R_a\psi(x) = e^{(1/2)a}\psi(e^a x) \qquad T_\beta\psi(x) = e^{i\beta x}\psi(x). \tag{10}$$

The infinitesimal operators are

$$I_\alpha = I_1 = {}^1/_2 + x\frac{d}{dx} \qquad I_\beta = I_2 = ix \tag{10a}$$

with the structure relation

$$[I_1, I_2] = I_2. \tag{10b}$$

Contraction with respect to the group of transformations $x' = e^{ax}$ leaves the group unchanged: the only non-vanishing structure constant is $C_{12}{}^2$ and (9) shows that this does not change. We can try, therefore, to transform I_1 and ϵI_2 with an ϵ dependent unitary matrix so that I_1 remain unchanged, the transformed ϵI_2 converge to I_2

$$S_\epsilon{}^{-1}I_1S_\epsilon = I_1 \qquad \epsilon S_\epsilon{}^{-1}I_2S_\epsilon \to I_2. \tag{11}$$

This is indeed possible: one has to choose $S_\epsilon = R_{\ln\epsilon}$. This commutes with I_1. It follows from the group relation that $S_\epsilon{}^{-1}T(\beta)S_\epsilon = R_{-\ln\epsilon}T(\beta)R_{\ln\epsilon} = T(\beta/\epsilon)$ and one has, hence, as $h \to 0$

$$\begin{aligned}\epsilon S_\epsilon{}^{-1}I_2S_\epsilon &= \epsilon S_\epsilon{}^{-1}\lim h^{-1}(T(h) - 1)S_\epsilon \\ &= \lim \epsilon h^{-1}(T(h/\epsilon) - 1) = I_2\end{aligned} \tag{11a}$$

Hence $\epsilon S_\epsilon{}^{-1}I_2S_\epsilon$ not only converges to $J_2 = I_2$ but remains equal to it for all ϵ.

It is more surprising, perhaps, to see that the same device is possible also if one contracts the group with respect to the subgroup of the $T(\beta)$. The contracted group is, in this case, the two parametric abelian group. We demand, in this case, that S_ϵ commute with I_β and hence, by (10a), that it be multiplication with a function of x. Because of S' unitary nature, we can give it the form $\exp(if(x, \epsilon))$. Transformation of the ϵI_α of (10a) with this gives

$$\begin{aligned}\epsilon S_\epsilon{}^{-1}I_\alpha S_\epsilon &= \epsilon e^{-if(x,\epsilon)}({}^1/_2 + x\,d/dx)e^{if(x,\epsilon)} \\ &= \epsilon({}^1/_2 + x\,d/dx) + \epsilon ix\,d/dx\,f(x, \epsilon).\end{aligned} \tag{12}$$

The first part of this converges to 0 as it should since (12) should converge to an operator which commutes with ix. The second part converges to $J_2 = ixf'(x) = ig(x)$ if one sets

$$f(x, \epsilon) = \epsilon^{-1}f(x). \tag{12a}$$

Hence, the transformations of the contracted group corresponding to the parameters α, β is multiplication with

$$e^{i\alpha g(x) + i\beta x} \tag{13}$$

which is again a faithful (reducible) representation of the contracted group.

The operators corresponding to finite group elements could have been obtained directly by transforming $O_{\alpha,\epsilon\beta} = T(\epsilon\beta)R_\alpha$ with $R_{\ln \epsilon}$ in the case of contraction with respect to R_α. Similarly, (13) could have been obtained directly by transforming $O_{\epsilon\alpha,\beta} = T(\beta)R_{\epsilon\alpha}$ with $\exp(\epsilon^{-1}if(x))$.

It is not clear how generally one can obtain a faithful representation of the contracted group as a limit of an ϵ dependent transform of a faithful representation of the original group and the substitution (6b) of its parameters. Certainly, the procedure is not applicable to irreducible representations of compact groups or, more generally, if the infinitesimal operators are bounded.

(*b*) *Representations of the Contracted Group from a Sequence of Representations.*—We shall now give a few examples for the second procedure, i.e., obtaining a representation of the contracted group by choosing a sequence of unitary representations $D^{(1)}, D^{(2)}, \ldots, D^{(l)}, \ldots$ so that each of the operators

$$I_{1\nu}^{(l)}, \epsilon I_{2\mu}^{(l)} \qquad (\nu = 1, 2, \ldots, r;\ \mu = 1, 2, \ldots, n - r) \tag{15}$$

converge to a finite operator as $\epsilon \to 0$ and $l \to \infty$. Alternately, we can ask that the transformation corresponding to finite group elements

$$D^{(l)}(b_{1\nu}, \epsilon b_{2\mu}) \tag{15a}$$

converge to a unitary representation of the contracted group as $\epsilon \to 0$, $l \to \infty$. The b are the parameters of the contracted group; the operator (15) corresponds, in the representation l, to the group element of the original group the parameters a of which are given by (6b). The formulation making use of the finite group elements (15a) is unambiguous because it deals with the convergence of unitary operators; the first one is usually easier to attack directly.

The convergence of the sequences of (15) and (15a) will depend not only on the values which ϵ assumes and on the corresponding representations $D^{(l)}$, it will also depend in which form that representation is assumed. Hence, method (*a*) can be considered as a special case of the present method in which all $D^{(l)}$ are unitary equivalent.

The contracted group is always an open group because the variability domain of the b is infinite. Hence its representations are, as a rule, infinite dimensional. If the $D^{(l)}$ are finite dimensional, they should be considered to affect only a finite number of the coordinates of Hilbert space.

The simplest non-commutative compact group is the three-dimensional rotation group. All its subgroups are one parametric, we shall contract it with respect to the rotations about the z axis of a rectangular coordinate

system of ordinary space. The contracted group is the Euclidean group of the plane, i.e., the inhomogeneous two-dimensional rotation group. We shall choose for $D^{(l)}$ the representation which is usually denoted[2] by $D^{(l)}$: it is $2l + 1$ dimensional (l being any integer) and is usually described in a space the coordinate axes of which are labeled with $m = -l, -l + 1, \ldots, l - 1, l$. Hence, we label the coordinate axes of Hilbert space with all integers m from $-\infty$ to ∞. In keeping with our previous notation, we call the infinitesimal element which corresponds to rotations about the z axis $M_z = I_1$. One can then write for $|m| \leq l$, $|m'| \leq l$

$$(I_1^{(l)})_{mm'} = (M_z^{(l)})_{mm'} = im\delta_{mm'} \tag{16}$$

$$(I_{2x}^{(l)})_{mm'} = (M_x^{(l)})_{mm'} = -\frac{1}{2}\sqrt{(l-m)(l+m')}\,\delta_{m'm+1} + \frac{1}{2}\sqrt{(l-m')(l+m)}\delta_{m'm-1}$$

$$(I_{2y}^{(l)})_{mm'} = (M_y^{(l)})_{mm'} = \frac{i}{2}\sqrt{(l-m)(l+m')}\,\delta_{m'm+1} + \frac{i}{2}\sqrt{(l-m')(l+m)}\,\delta_{m'm-1}.$$

All matrix elements vanish if either $|m|$ or $|m'|$ is larger than l. As $l \to \infty$, the I_1 converge to a definite operator

$$(J_1)_{mm'} = im\,\delta_{mm'}. \qquad (-\infty < m, m' < \infty) \tag{17a}$$

In fact, the convergence is strong in the sense that $I_1\varphi$ converges strongly to $J_1\varphi$ if φ is in the definition domain of J_1. If ϵ and l converge to zero and infinity, respectively, in such a way that $l\epsilon \to \Xi$, the other two infinitesimal elements will also converge to J_{2x}, J_{2y}, respectively, where

$$(J_{2x})_{mm'} = \frac{1}{2}\Xi(\delta_{m'm-1} - \delta_{m'm+1}) \tag{17b}$$

$$(J_{2y})_{mm'} = \frac{i}{2}\Xi\,(\delta_{m'm-1} + \delta_{m'm+1}).$$

It follows from the fact that there are only a finite number (two) nonvanishing matrix elements in both I_{2x} and I_{2y} that the operators satisfy the commutation relations of the contracted group

$$[J_{2x}, J_{2y}] = 0 \qquad [J_1, J_{2x}] = J_{2y} \qquad [J_1, J_{2y}] = -J_{2x}. \tag{18}$$

It further follows from

$$(I_1^{(l)})^2 + (I_{2x}^{(l)})^2 + (I_{2y}^{(l)})^2 = -l(l+1)$$

by multiplication with ϵ^2 and going to the limit in the above way that

$$J_{2x}^2 + J_{2y}^2 = -\Xi^2. \tag{18a}$$

It was important, for obtaining convergent sequences of $I_1^{(l)}$ and $\epsilon I_2^{(l)}$, to have assumed the $D^{(l)}$ in the form given by (16). This form was reduced out with respect to the subgroup S and this caused, in this case, the convergence of the sequence $I_1^{(l)}$. The convergence of the $\epsilon I_2^{(l)}$ does not actually follow from the convergence of the $I_1^{(l)}$, and hence from the reduced out form of the representations $D^{(l)}$, but was made at least possible by this circumstance.

Before going over to the investigation of the Lorentz groups, it may be worth while to make a final remark about the above contraction, even though it has little to do with our subject. We shall determine, first, the matrices which correspond to finite group elements. For this purpose, it is useful to consider that form of the above representation in which the Hilbert space consists of functions of x and y and the infinitesimal operators have the natural form for infinitesimal operators of the Euclidean group (α and r are polar coordinates $x = r \cos \alpha$, $y = r \sin \alpha$)

$$J_1' = -\partial/\partial\alpha = y\partial/\partial x - x\partial/\partial y \tag{19}$$

$$J_{2x}' = \partial/\partial y \qquad J_{2y}' = -\partial/\partial x. \tag{19a}$$

One should keep in mind that J_{2x}' arose from I_{2x} which is the infinitesimal rotation about the x axis and corresponds to a displacement in the $-y$ direction. Similarly J_{2y}' corresponds to a displacement in the x direction.

The function $\varphi(\alpha, r)$ corresponds in the new Hilbert space to the vector which has the components φ_m in the Hilbert space of (17a), (17b). It further follows from (18a) that

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\varphi = -\Xi^2 \varphi \tag{20}$$

whence one can write

$$\begin{aligned} \varphi(x, y) &= \int e^{-i\Xi(x \cos \alpha' + y \sin \alpha')} g(\alpha') d\alpha' \\ &= \int e^{-i\Xi r \cos (\alpha - \alpha')} g(\alpha') d\alpha' \\ &= \int e^{-i\Xi r \cos \alpha'} g(\alpha - \alpha') d\alpha'. \end{aligned} \tag{20a}$$

All integrations are from 0 to 2π. Expanding $g(\alpha - \alpha')$ into a Fourier series of $\alpha - \alpha'$, one finds

$$\varphi(x, y) = \int e^{-i\Xi r \cos \alpha'} \sum g_m e^{-im(\alpha - \alpha')} d\alpha' \tag{20b}$$

m runs from $-\infty$ to ∞. The last form makes it easy to calculate J_1' by (19), the first form permits one to calculate J_{2x}' and J_{2y}' easily. Comparison of the expressions obtained in this way with (17a) and (17b) shows that $g_m = \varphi_m$. In order to remain in keeping with the usual notation, we define[3]

$$2\pi J_m(z) = i^m \int e^{-iz\cos\alpha'} e^{im\alpha'} d\alpha'.$$

J_m is then the ordinary Bessel function of order m. This permits one to write for (20b)

$$\varphi(\alpha, r) = 2\pi \sum \varphi_m i^{-m} e^{-im\alpha} J_m(\Xi r). \tag{21}$$

Because of (19a), one can write down at once the finite group operations. In particular, for the displacement $T(\xi, \eta)$ by ξ and η in the x and y directions one has

$$T(\xi, \eta)\varphi(x, y) = \varphi(x - \xi, y - \eta). \tag{22}$$

Hence, denoting the matrix for the same operation in the original Hilbert space by $T(\xi, \eta)_{mm'}$, one has

$$\varphi(x - \xi, y - \eta) = 2\pi \sum_m \sum_{m'} T(\xi, \eta)_{mm'} \varphi_{m'} i^{-m} e^{-ima} J_m(\Xi r). \tag{22a}$$

This permits an explicit determination of the $T(\xi, \eta)_{mm'}$. We shall not carry this out completely but set only $r = 0$ in (22a). Since all $J_m(0) = 0$ except $J_0(0) = 1$, the summation over m disappears on the right side. The left side becomes, at the same time by (21)

$$\varphi(-\xi, -\eta) = 2\pi \sum \varphi_m i^{-m} e^{-im(\beta + \pi)} J_m(\Xi\rho) \tag{22b}$$

where β, ρ are the polar coordinates for ξ, η. Comparing (22a) and (22b) one finds, with $\xi = 0$, $\beta = {}^1/_2\pi$, $\eta = \rho$

$$J_m(\Xi\rho) = T(0, \rho)_{0m}. \tag{23}$$

The group relations and the form of the infinitesimal operators (19a) gives at once the most important relations for Bessel functions, such as the addition theorem, differential equation (cf. (20)), etc. Up to this point the argument is not new but merely a repetition, for the two dimensional Euclidean group, of a similar reasoning given before[4] for the rotation group. This led to the equation[4]

$$D^{(l)}(0, \beta 0)_{0m} = \bar{P}^m(\cos\beta) = \left(\frac{(l - m)!}{(l + m)!}\right)^{1/2} P_l^m(\cos\beta) \tag{24}$$

in which P_l^m is Legendre's associated function of the first kind, $\bar{P}_l^m$ is normalized to the same value as $P_l^0 = P_l$. Furthermore, 0, β, 0 is the rotation about the x axis so that, by the definition of $T(0, \rho)$

$$T(0, \rho)_{m'm} = (-)^{m'} \lim D^{(l)} \qquad (0, \Xi\rho/l, 0)_{m'm}. \tag{25}$$

This, together with (23), gives the asymptotic expression for the associated Legendre functions[5]

$$\lim_{l \to \infty} \bar{P}_l^m(\cos(\rho/l)) = J_m(\rho). \tag{25a}$$

III. Contraction of Lorentz Groups

Let us consider, first, the inhomogeneous Lorentz group with one space-like, one time-like dimension. It is given by the transformations

$$\begin{aligned} x' &= x \cosh \lambda + t \sinh \lambda + a_x \\ t' &= x \sinh \lambda + t \cosh \lambda + a_t. \end{aligned} \tag{26}$$

We wish to contract it with respect to the subgroup of time displacements $t' = t + a_t$. The infinitesimal elements of (26) are: time displacement I_1, space displacement I_{2x} and "rotation" in space-time $I_{2\lambda}$. Their commutation relations read

$$[I_1, I_{2x}] = 0 \qquad [I_1, I_{2\lambda}] = -I_{2x} \qquad [I_{2x}, I_{2\lambda}] = -I_1. \tag{27}$$

Hence, by (9), the commutation relations of the contracted group are

$$[J_1, J_{2x}] = 0 \qquad [J_1, J_{2\lambda}] = -J_{2x} \qquad [J_{2x}, J_{2\lambda}] = 0. \tag{27a}$$

The "rotations" in space-time, together with the displacements in space, form a commutative invariant subgroup.

The matrices

$$\begin{Vmatrix} \cosh \lambda & \sinh \lambda & a_x \\ \sinh \lambda & \cosh \lambda & a_t \\ 0 & 0 & 1 \end{Vmatrix} \tag{26a}$$

form a natural, though not unitary, representation of the group of transformations (26). We can carry out the contraction by setting $a_t = b_t$, $\lambda = \epsilon v$, $a_x = \epsilon b_x$ or $\lambda = v/c$, $a_x = b_x/c$ and letting ϵ converge to 0, or c converge to infinity. If we do this directly in (26a), the representation will not remain faithful for the contracted group. We shall transform therefore (26a) with a suitable ϵ (or c) dependent matrix: multiply the first row with c, the first column with $1/c$. If c goes to infinity in the matrix obtained in this way, one obtains the transformations of the contracted group

$$\begin{aligned} x' &= x + vt + b_x \\ t' &= t + b_t. \end{aligned} \tag{27a}$$

It is the inhomogeneous Galilei group with one spatial dimension. The transformations $x' = x + vt + b_x$, $t' = t$ form the commutative invariant subgroup.

The same contraction can be carried out for an inhomogeneous Lorentz group with an arbitrary number of spatial dimensions. The only difference is that the subgroup S, with respect to which the contraction is carried out, contains not only the displacements in time as in the above example, but also all purely spatial rotations, i.e., all homogeneous transformations

which leave t invariant. The invariant subgroup of the contracted group consists of all spatial displacements and Galilei transformations: $x_i' = x_i' + v_i t + b_i$, $t' = t$.

Contraction of the Unitary Representations of the Lorentz Groups.—We shall be principally concerned here with the group of the special theory of relativity, i.e., the inhomogeneous Lorentz group with three space-like and one time-like dimension. The subgroup S with respect to which we shall contract it contains the displacements in time, the spatial rotations, and the products of these operations. The contracted group is the ordinary Galilei group, i.e., the group of classical mechanics.

We shall denote the displacement operators in the direction of the three space-like axes by I_k ($k = 1, 2, 3$), the displacement in the direction of the time axis by I_0. The rotations in the kl plane will be denoted by I_{kl}, the acceleration in the direction of the k axis by I_{k0}. I_0 and I_{kl} span the subgroup S.

The quantity

$$I_1^2 + I_2^2 + I_3^2 - I_0^2 = P \tag{28}$$

is a constant in every irreducible representation and the irreducible representations can be divided into three classes according to the value of this constant.[6] In the first class, $P < 0$ and the momenta (which are $-i$ times the infinitesimal operators) are space-like. In the second class $P = 0$ and the momenta form a null vector, $P > 0$ in the third class.

It is generally admitted that the representations of the first class have no physical significance because the momenta of all observed particles are time-like or null vectors. Hence we shall investigate only the simplest one of these representations. Its operators are most easily given in the Hilbert space of functions of three variables p_1, p_2, p_3 which are restricted to the outside of a sphere of radius $\sqrt{-P}$. The expressions for the infinitesimal operators are

$$I_k = ip_k \qquad I_0 = -i(p_1^2 + p_2^2 + p_3^2 + P)^{1/2}. \tag{28a}$$

This last equation also shows the reason for the variables p_k to be restricted by $p_1^2 + p_2^2 + p_3^2 > -P$: if this inequality is not fulfilled, I_0 ceases to be skew hermitean and, hence, the representation is not unitary. Further

$$I_{kl} = p_l \partial/\partial p_k - p_k \partial/\partial p_l \tag{28b}$$

and

$$I_{k0} = -(p_1^2 + p_2^2 + p_3^2 + P)^{1/2} \partial/\partial p_k. \tag{28c}$$

It is useful to introduce new variables instead of the p_1 both in order to simplify the definition domain of the variables and also to bring the operations of the subgroup S into a form which is independent of P. This can

be done, most simply, by introducing as new variables

$$p_0 = (p_1^2 + p_2^2 + p_3^2 + P)^{1/2} \qquad \Omega_k = p_k/(p_0^2 - P)^{1/2}. \tag{29}$$

The Ω_k are restricted to a unit sphere, p_0 changes from 0 to ∞. In these variables, the scalar product between φ and ψ reduces to

$$(\varphi, \psi) = \int d\Omega \int dp_0 (p_0^2 - P)^{1/2} \bar{\varphi}\psi \tag{29a}$$

$d\Omega$ being the surface element of the unit sphere over which the integration with respect to the Ω is to be extended. In terms of the new variables, the infinitesimal operators assume the form

$$I_{kl} = \Omega_l \partial/\partial\Omega_k - \Omega_k \partial/\partial\Omega_l \qquad I_0 = -ip_0 \tag{30}$$

and

$$I_k = i\Omega_k (p_0^2 - P)^{1/2} \tag{30a}$$

$$I_{k0} = -(p_0^2 - P)^{1/2}\Omega_k \partial/\partial p_0 - \frac{p_0}{(p_0^2 - P)^{1/2}} \sum_l \Omega_l (\Omega_l \partial/\partial\Omega_k - \Omega_k \partial/\partial\Omega_l).$$

If we now set $J_k = \epsilon I_k$, the J_k will converge to zero unless $-P$ becomes inversely proportional to ϵ^2, i.e., unless $-\epsilon^2 P$ converges to a definite limit P. If this is assumed, the second term of ϵI_{k0} will converge to zero and the infinitesimal elements of the representation of the contracted group become

$$J_{kl} = \Omega_l \partial/\partial\Omega_k - \Omega_k \partial/\partial\Omega_l \qquad J_0 = -ip_0 \tag{31}$$

$$J_{k0} = -\mathrm{P}\Omega_k \partial/\partial p_0 \qquad J_k = i\,\mathrm{P}\Omega_k. \tag{31a}$$

These operators indeed span a unitary representation of the Galilei group: they correspond to case II with $m = 0$ of a recent determination of these representations.[7] The p_k of this article correspond to our $\mathrm{P}\Omega_k$, the variable s is given by $i\,\mathrm{P}\,\partial/\partial p_0$. One also understands now why it was impossible to find a physical interpretation to this representation: it is the limiting case of a representation of the relativistic group with imaginary mass $P^{1/2}$. The same is probably true of the other true representations[7] of the Galilei group.

We go over now to the investigation of the simplest representation with positive P, i.e., the representation of the Klein-Gordon equation. The infinitesimal operators are again given by (28a), (28b), (28c). However, since P is positive, the variability domain of the p extends over the whole three dimensional space. The scalar product of two functions φ and ψ is now given by

$$(\varphi, \psi) = \int\int\int dp_1\, dp_2\, dp_3\, p_0^{-1}\, \bar{\varphi}\psi \tag{32}$$

in which p_0 is still given by (29).

If we set, in the sense of (6),

$$J_k = \epsilon I_k = \epsilon i p_k \tag{33}$$

the operation of time displacement becomes

$$iJ_0 = iI_0 = (-I_1^2 - I_2^2 - I_3^2 + P)^{1/2} = \epsilon^{-1}(-J_1^2 - J_2^2 - J_3^2 + \epsilon^2 P)^{1/2} \tag{33a}$$

Since the J_k are skew hermitean, their squares are positive definite hermitean operators. Since $\epsilon^2 P$ is also positive the expectation value $(\varphi, iJ_0\varphi)$ of iJ_0 is, for any state φ, greater than ϵ^{-1} times the expectation value of iJ_1. It follows that if $J_1\varphi$ converges to a vector in Hilbert space as $\epsilon \to 0$, the vector $J_0\varphi$ must grow beyond all limits. The same is true, of course, for the other J_k. It follows that the representations considered cannot be contracted in the sense discussed in the previous sections and the same is true of all representations of the classes $P \geq 0$.

It is possible, however, to contract these representations to representations up to a factor of the Galilei group. The commutation relations of the infinitesimal elements of representations up to a factor differ from the commutation relations of real representations by the appearance of a constant in the structural relations. Hence

$$[J_\alpha, J_\beta] = \sum c_{\alpha\beta}^\gamma J_\gamma + b_{\alpha\beta} 1 \tag{34}$$

where $c_{\alpha\beta}^\gamma$ are the structure constants of the group to be represented (in our case, the inhomogeneous Galilei group) and the $b_{\alpha\beta}1$ are multiples of the unit operator. One will, therefore, obtain infinitesimal elements of representations up to a factor if one sets, instead of (6)

$$J_{1\nu} = I_{1\nu} - a_{1\nu}1 \qquad J_{2\nu} = \epsilon I_{2\nu} - a_{2\nu}1 \tag{34a}$$

in which all a may depend on ϵ. Since the additional terms in (34a) commute with all other operators, these additional terms will not affect the left side of (34). Hence, they must be compensated also on the right side and this is done by the additional terms $b_{\alpha\beta}1$. The point of introducing the terms $a1$ in (34a), which then necessitates the introduction of the b in (34), is that the right sides of (34a) may converge to finite non-vanishing operators even if the $I_{1\nu}$, $\epsilon I_{2\nu}$ cannot be made to converge.

The above generalization of the concept of contraction indeed allows a contraction of the representations given by (28a), (28b), (28c) also for $P > 0$. As we let P go to infinity, I_0 will also tend to infinity ($(I_0\varphi, I_0\varphi)$ converges to infinity for all φ). However, subtracting $-iP^{1/2}1$ from I_0, it will converge to

$$J_0 = \lim -i(-I_1^2 - I_2^2 - I_3^2 + P)^{1/2} + iP^{1/2} = (i/2P^{1/2})(I_1^2 + I_2^2 + I_3^2) = (i/2P^{1/2}\epsilon^2)(J_1^2 + J_2^2 + J_3^2). \tag{35}$$

This shows that J_0 will converge to a finite operator if the $J_k = \epsilon I_k$ do and if $P^{1/2}\epsilon^2$ converges to a finite constant m as $P \to \infty$, $\epsilon \to 0$. Both can be accomplished by assuming the representation in such a form that, instead of (28)

$$I_k = ip_k/\epsilon \qquad I_0 = -i(P + (p_1^2 + p_2^2 + p_3^2)/\epsilon^2)^{1/2}. \tag{36}$$

This is indeed possible because the variability domain of the p_k is unrestricted and the above form of the infinitesimal elements can be obtained by unitary transformation of the operators given in (28a). Such a transformation leaves the I_{kl} of (28b) unchanged but transforms the I_{k0} of (28c) into

$$I_{k0} = -\epsilon(P + (p_1^2 + p_2^2 + p_3^2)/\epsilon^2)^{1/2}\partial/\partial p_k. \tag{36b}$$

Hence we shall have

$$\begin{aligned} J_0 &= \lim -i(P + (p_1^2 + p_2^2 + p_3^2)/\epsilon^2)^{1/2} + iP^{1/2} = \\ &\qquad -(i/2m)(p_1^2 + p_2^2 + p_3^2) \\ J_{kl} &= p_l\partial/\partial p_k - p_k\partial/\partial p_l \qquad (37.1) \\ J_k &= \lim \epsilon(ip_k/\epsilon) = ip_k \\ J_{k0} &= \lim -\epsilon(P + (p_1^2 + p_2^2 + p_3^2)/\epsilon^2)^{1/2}\epsilon\partial/\partial p_k \\ &= \lim -(P\epsilon^4 + \epsilon^2(p_1^2 + p_2^2 + p_3^2))^{1/2}\partial/\partial p_k = -m\partial/\partial p_k. \qquad (37.2) \end{aligned}$$

The reader familiar with the transition from the Klein-Gordon to the Schrödinger equation will recognize the increase of the rest mass with increasing c and the elimination of this rest mass by the subtraction of $-iP^{1/2}1$ from the infinitesimal operator of the time-displacement operator. The infinitesimal operators (37.1), (37.2) for the contracted group are in fact those of Schrödinger's theory. It is likely that a similar contraction is possible also for the other representations with positive rest mass (i.e., $P > 0$) but this and the behavior of the representations with $P = 0$ will not be further discussed here.

[1] Segal, I. E., *Duke Math. J.*, **18,** 221 (1951).

[2] Cf. e.g. Wigner, E., *Gruppentheorie und ihre Auwendiengen* etc., Friedr. Vieweg, Braunschweig (1931) and Edwards Brothers, Ann Arbor (1944). Chapter XV.

[3] Cf. Jahnke, E., and Emde, F., *Tables of Functions*, Dover Publications, 1943, p. 149; or Watson, G. N., *Treatise on Bessel Functions*, Cambridge Univ. Press, 1922, p. 19ff.

[4] Reference 2, Chapter XIX, particularly p. 230, 232. Cf. also Wigner, E. P., *J. Franklin Inst.*, **250,** 477 (1950), and Godement, R., *Trans. Amer. Math. Soc.* **73,** 496 (1952).

[5] For $m = 0$ this is given on p. 65, of Watson's *Bessel Functions* (ref. 3).

[6] Cf. e.g. Bargmann, V., and Wigner, E. P., these PROCEEDINGS, **34,** 211 (1948) and further literature quoted there.

[7] Inonu, E., and Wigner, E. P., *Nuovo cimento*, **9,** 705 (1952).

ERGODIC PROPERTY OF THE BROWNIAN MOTION PROCESS

BY G. KALLIANPUR AND H. ROBBINS*

THE INSTITUTE FOR ADVANCED STUDY AND UNIVERSITY OF NORTH CAROLINA

Communicated by J. von Neumann, April 16, 1953

1. Introduction.—In a previous paper ("On the Equidistribution of Sums of Independent Random Variables," to appear elsewhere; an abstract will appear in *Bull. Am. Math. Soc.*, **59,** (May, 1953); we shall refer to this paper as [1]) we considered some properties of the sequence S_n of partial sums of independent and identically distributed random variables or random vectors in two dimensions. Here we show in Theorems 1–4 that the results of [1] carry over to the Brownian motion process in one and two dimensions.

2. The Two-Dimensional Case.—Let $X(t)$ denote the Brownian motion (Wiener) process on the line: $X(0) \equiv 0$, $X(t)$ is continuous for all t with probability 1, and for any $t_0 < t_1 < \ldots < t_n$ the random variables $X(t_j) - X(t_{j-1})$, $j = 1, \ldots, n$, are independent and normally distributed with zero means and variances $t_j - t_{j-1}$. Let $V(t) = (X(t), Y(t))$ denote the Brownian motion process in the plane, the two components of $V(t)$ being independent one-dimensional Brownian motion processes. Suppose that $f(x, y)$, $g(x, y)$ are real valued functions which are bounded and summable in the plane $-\infty < x < \infty$, $-\infty < y < \infty$, and set $\bar{f} = \int\int f(x, y)\, dxdy$, $\bar{g} = \int\int g(x, y)\, dxdy$, where here and in the sequel an integral sign without limits denotes integration over $(-\infty, \infty)$. We shall prove the following two theorems for plane Brownian motion. The corresponding results for the one-dimensional case involve no essentially new arguments and will be stated without proof at the end of the paper.

THEOREM 1. *If $\bar{f} \neq 0$ then for every u,*

$$\lim_{T \to \infty} \Pr\left[\frac{2\pi}{\bar{f} \log T} \int_0^T f(V(t))\, dt \leq u\right] = G(u), \qquad (2.1)$$

where $G(u) = 1 - e^{-u}$ for $u \geq 0$, $= 0$ for $u < 0$.

THEOREM 2. *If $\bar{g} \neq 0$ then*

$$\lim_{T \to \infty} \frac{\int_0^T f(V(t))\, dt}{\int_0^T g(V(t))\, dt} = \frac{\bar{f}}{\bar{g}} \text{ in probability.} \qquad (2.2)$$

Proofs of Theorems 1 and 2: Assume $\bar{f} \neq 0$ and $\bar{g} \neq 0$ and define

$$Z_n(f) = \frac{2\pi}{\bar{f} \log n} \int_0^n f(V(t))\, dt, \qquad W_n(f) = \frac{2\pi}{\bar{f} \log n} \sum_{j=1}^{n} f(V(j)),$$

with corresponding definitions for $Z_n(g)$ and $W_n(g)$. For any positive

integer n, each component of $V(n)$ is the sum of n independent random variables, each distributed normally with zero mean and unit variance, since

$$X(n) = \sum_{j=1}^{n} [X(j) - X(j-1)], \qquad Y(n) = \sum_{j=1}^{n} [Y(j) - Y(j-1)].$$

A theorem in [1] gives

$$\lim_{n\to\infty} \Pr[W_n(f) \leq u] = G(u). \tag{2.3}$$

We shall later prove as Lemma 1 that

$$\lim_{n\to\infty} E[Z_n(f) - W_n(f)]^2 = 0. \tag{2.4}$$

It follows from (2.4) that $Z_n(f) - W_n(f)$ tends to zero in probability, and hence from (2.3) that $Z_n(f)$ has the same limiting distribution, $G(u)$, as $W_n(f)$. This proves (2.1) as $T \to \infty$ through integer values, and the extention to arbitrary T is immediate, proving Theorem 1.

To prove Theorem 2 we shall later prove as Lemma 2 that

$$\lim_{n\to\infty} E[Z_n(f)Z_n(g) - W_n(f)W_n(g)] = 0, \tag{2.5}$$

and we make use of the fact, proved in [1], that

$$\lim_{n\to\infty} EW_n^2(f) = \lim_{n\to\infty} EW_n^2(g) = \lim_{n\to\infty} EW_n(f)W_n(g) = \int_0^\infty u^2 dG(u) = m_2, \text{ say.} \tag{2.6}$$

From (2.4)–(2.6) it follows that

$$\lim_{n\to\infty} EZ_n^2(f) = \lim_{n\to\infty} EZ_n^2(g) = \lim_{n\to\infty} EZ_n(f)Z_n(g) = m_2,$$

$$\text{and hence } \lim_{n\to\infty} E[Z_n(f) - Z_n(g)]^2 = 0. \tag{2.7}$$

As in [1], given $\epsilon > 0$ we choose $\delta = \delta(\epsilon) > 0$ such that $\delta \leq \epsilon$, $G(\delta) \leq {}^1/_2\epsilon$, and $N = N(\epsilon)$ (by (2.1) with f replaced by g) such that

$$n \geq N \text{ implies } \Pr[Z_n(g) > \delta] \geq 1 - G(\delta) - {}^1/_2\epsilon \geq 1 - \epsilon.$$

Since $\Pr[\{Z_n(f) - Z_n(g)\}^2 < \delta^3] \geq 1 - \dfrac{E[Z_n(f) - Z_n(g)]^2}{\delta^3}$,

if we choose (by (2.7)) $K = K(\epsilon)$ such that

$$n \geq K \text{ implies } E[Z_n(f) - Z_n(g)]^2 \leq \delta^4,$$

then if $n \geq$ max. (N, K),

$$\Pr[Z_n(g) > \delta] \geq 1 - \epsilon, \qquad \Pr[\{Z_n(f) - Z_n(g)\}^2 < \delta^3] \geq 1 - \epsilon.$$

which in turn imply that

$$\Pr\left[\left\{\frac{Z_n(f)}{Z_n(g)} - 1\right\}^2 < \epsilon\right] \geq \Pr[Z_n(g) > \delta,\ |Z_n(f) - Z_n(g)|^2 < \delta^3] \geq 1 - 2\epsilon.$$

Since ϵ was arbitrary this proves (2.2) as $T \to \infty$ through integer values. Again, the extension to arbitrary T is immediate. Finally, the restriction that $\hat{f} \neq 0$ can be dropped by a simple argument and the proof of Theorem 2 is complete.

3. *Proof of the Lemmas.*—We shall prove Lemma 2 first. We have

$$Z_n(f) = \frac{2\pi}{\hat{f} \log n} \sum_{j=1}^{n} \int_{j-1}^{j} f(V(t))\, dt,\quad W_n(f) = \frac{2\pi}{\hat{f} \log n} \sum_{j=1}^{n} \int_{j-1}^{j} f(V(j))\, dt.$$

Set $D_n = E[Z_n(f)Z_n(g) - W_n(f)W_n(g)]$

$$= \frac{4\pi^2}{\hat{f}\hat{g}\,(\log n)^2} \sum_{j,\,k=1}^{n} \int_{j-1}^{j}\int_{k-1}^{k} E[f(V(t))g(V(u)) - f(V(j))g(V(k))]\, dudt$$

$$= \frac{4\pi^2}{\hat{f}\hat{g}\,(\log n)^2} \sum_{j,\,k=1}^{n} a_{jk},$$

where

$$a_{jk} = \iint_{R_{jk}} [\theta(t, u) - \theta(j, k)]\, dudt, \qquad R_{jk} = \{j - 1 < t < j;\ k - 1 < u < k\},$$

$$\theta(t, u) = E[f(V(t))g(V(u))].$$

We want to show that $D_n \to 0$ as $n \to \infty$, i.e., that

$$\lim_{n \to \infty} (\log n)^{-2} \sum_{j,\,k=1}^{n} a_{jk} = 0. \tag{3.1}$$

To evaluate $\theta(t, u)$ we observe that for $t_1 < t_2$ the random vector $V(t_2) - V(t_1) = (X(t_2) - X(t_1),\ Y(t_2) - Y(t_1))$ has the joint probability density

$$\frac{1}{2\pi(t_2 - t_1)} e^{-\frac{x^2 + y^2}{2(t_2 - t_1)}}.$$

Hence for $0 < t < u$,

$$\theta(t, u) = \frac{1}{4\pi^2 t(u - t)} \int\int\int\int f(x, y)g(x + x', y + y') \, e^{-\frac{x^2 + y^2}{2t} - \frac{x'^2 + y'^2}{2(u - t)}}\, dxdydx'dy'$$

$$= \frac{1}{4\pi^2 t(u - t)} \int\int\int\int f(x, y)g(\zeta, \eta) e^{-Q(t, u)}\, dxdyd\zeta d\eta,$$

where we have set

$$Q(t, u) = \frac{1}{2}\left\{\frac{x^2 + y^2}{t} + \frac{(x - \zeta)^2 + (y - \eta)^2}{u - t}\right\}.$$

In what follows C will denote any constant whose numerical value is immaterial, and $F = \sup|f(x, y)|$, $G = \sup|g(x, y)|$. Then if $(t, u)\epsilon R_{jj}$ or $\epsilon R_{j, j+1}(j \geq 2)$,

$$|\theta(t, u)| \leq \frac{G}{4\pi^2 t(u - t)} \int\int |f(x, y)| e^{-\frac{x^2 + y^2}{2t}} dxdy \cdot$$

$$\int\int e^{-\frac{x'^2 + y'^2}{2(u - t)}} dx'dy' \leq \frac{G}{2\pi t} \int\int |f(x y)| dxdy \leq \frac{C}{j - 1}. \quad (3.2)$$

Also, for $0 < t < u$,

$$|\theta(t, u)| \leq \frac{1}{4\pi^2 t(u - t)} \int\int\int\int |f(x, y) \cdot g(\zeta, \eta)|$$

$$e^{-\frac{x^2 + y^2}{2t}} dxdyd\zeta d\eta \leq \frac{F}{2\pi(u - t)} \int\int |g(\zeta, \eta)| d\zeta d\eta \leq \frac{C}{u - t}. \quad (3.3)$$

We write $\sum_{j, k = 1}^{n} a_{jk}$ as the sum of the following terms:

(a) $\sum_{j = 2}^{n} a_{jj}$, (b) $\sum_{k = 3}^{n} a_{1k}$, (b′) $\sum_{i = 3}^{n} a_{j1}$, (c) $\sum_{j = 2}^{n - 1} a_{j, j+1}$,

(c′) $\sum_{k = 2}^{n - 1} a_{k+1, k}$, (d) $\sum_{2 \leq j \leq k - 2 \leq n - 2} a_{jk}$, (d′) $\sum_{2 \leq k \leq j - 2 \leq n - 2} a_{jk}$, and

(e) $a_{11} + a_{12} + a_{21}$.

To prove (3.1) it will suffice to show that each of (a), (b), (c), (d) is $o\{(\log n)^2\}$.

(a) From (3.2), if $(t, u)\epsilon R_{jj}$, $j \geq 2$, $|\theta(t, u)| \leq \frac{C}{j - 1}$.

Also, $|\theta(j, j)| = \frac{1}{2\pi j}\left|\int\int f(x, y)g(x, y)e^{-\frac{x^2 + \cdots}{2j}} dxdy\right| \leq \frac{C}{j - 1}$.

Hence $\left|\sum_{j = 2}^{n} a_{jj}\right| \leq C \cdot \sum_{j = 2}^{n} \frac{1}{j - 1} = 0(\log n)$. (3.4)

(b) From (3.3), if $(t, u)\epsilon R_{1k}$, $k \geq 3$,

$$|\theta(t, u)| \leq \frac{C}{u - t} \leq \frac{C}{k - 2}.$$

$$\text{Also, } |\theta(1, k)| \leq \frac{C}{k - 2}.$$

$$\text{Hence } \left| \sum_{k=3}^{n} a_{1k} \right| \leq C \cdot \sum_{k=3}^{n} \frac{1}{k-2} = 0(\log n). \tag{3.5}$$

(c) Again from (3.2),

$$\sum_{j=2}^{n-1} a_{j,\,j+1} = 0(\log n). \tag{3.6}$$

(d) If $(t, u) \in R_{jk}$, $2 \leq j \leq k-2 \leq n-2$, then $Q(t, u) \geq 0$ and

$$\theta(t, u) - \theta(j, k) = \frac{1}{4\pi^2 t(u-t)} \int \cdots \int f(x, y) g(\zeta, \eta) \left[e^{-Q(t,u)} - e^{-Q(j,k)} \right] dx \ldots d\eta + \frac{1}{4\pi^2} \left[\frac{1}{t(u-t)} - \frac{1}{j(k-j)} \right] \int \cdots \int \times f(x,y) g(\zeta, \eta) e^{-Q(j,k)}\, dx \ldots d\eta = J_1 + J_2.$$

Setting
$R_a = \{|x| \leq a, |y| \leq a, |\zeta| \leq a, |\eta| \leq a\}$, R_a' = complement of R_a, let $\epsilon > 0$ be arbitrary and choose $a > 0$ such that

$$\int \cdots \int_{R_a'} \left| f(x, y) g(\zeta, \eta) \right| dx \ldots d\eta < \epsilon.$$

We can write

$$J_1 = \frac{1}{4\pi^2 t(u-t)} \int \cdots \int_{R_a'} + \frac{1}{4\pi^2 t(u-t)} \int \cdots \int_{R_a} = J_1' + J_1'',$$

where

$$|J_1'| \leq \frac{\epsilon}{4\pi^2 t(u-t)} \leq \frac{C\epsilon}{(j-1)(k-j-1)}. \tag{3.7}$$

To obtain a bound for J_1'' we observe that

$$Q(t, u) - Q(j, k) = \frac{x^2+y^2}{2}\left(\frac{1}{t} - \frac{1}{j}\right) + \frac{(x-\zeta)^2 + (y-\eta)^2}{2}\left(\frac{1}{u-t} - \frac{1}{k-j}\right) \leq \frac{x^2+y^2}{2}\left(\frac{1}{j-1} - \frac{1}{j}\right) + \frac{(x-\zeta)^2 + (y-\eta)^2}{2}\left(\frac{1}{k-j-1} - \frac{1}{k-j}\right) = \frac{x^2+y^2}{2(j-1)j} + \frac{(x-\zeta)^2 + (y-\eta)^2}{2(k-j-1)(k-j)} = A (\geq 0), \text{ say.}$$

Also,

$$Q(t, u) - Q(j, k) \geq \frac{(x-\zeta)^2 + (y-\eta)^2}{2}\left(\frac{1}{k-j+1} - \frac{1}{k-j}\right) = -\frac{(x-\zeta)^2 + (y-\eta)^2}{2(k-j)(k-j+1)} \geq -A.$$

Hence

$$e^{-A} - 1 \leq e^{-\{Q(t,u)-Q(j,k)\}} - 1 \leq e^{A} - 1, \quad \text{and}$$

$$\left| e^{-\{Q(t,u)-Q(j,k)\}} - 1 \right| \leq \max.\ [e^{A} - 1,\ 1 - e^{-A}] = e^{A} - 1.$$

Therefore

$$|J_1''| \leq \frac{1}{4\pi^2 t(u-t)} \int_{R_a} \cdots \int |f(x,y)g(\xi,\eta)| \cdot e^{-Q(j,k)} \cdot \left| e^{-\{Q(t,u)-Q(j,k)\}} - 1 \right| dx \ldots d\eta \leq \frac{FG}{4\pi^2 (j-1)(k-j-1)} \int_{R_a} \cdots \int (e^{A} - 1)\, dx \ldots d\eta$$

Since $e^x \leq 1 + xe^b$ for $0 \leq x \leq b$, we have in R_a

$$e^{A} - 1 \leq Ae^{\left[\frac{a^2}{(j-1)j} + \frac{4a^2}{(k-j-1)(k-j)}\right]} \leq Ae^{\left[\frac{a^2}{2} + 2a^2\right]},$$

and hence

$$|J_1''| \leq \frac{C(a)}{(j-1)(k-j-1)} \left[\frac{1}{(j-1)j} + \frac{1}{(k-j-1)(k-j)}\right]. \quad (3.8)$$

Turning to J_2, we have

$$|J_2| \leq C \left| \frac{1}{t(u-t)} - \frac{1}{j(k-j)} \right| = C \left| \frac{j(k-j) - t(u-t)}{t(u-t)j(k-j)} \right| \leq C \frac{\left| j(k-j) - t(u-t) \right|}{(j-1)j(k-j-1)(k-j)}.$$

Since $1 - 2j = j(k-j) - jk + (j-1)^2 \leq j(k-j) - t(u-t) \leq j(k-j) - (j-1)(k-1) + j^2 = k + j - 1$,

$$\left| j(k-j) - t(u-t) \right| \leq \max.\ [k + j - 1,\ 2j - 1] = k + j - 1 = (k-j-1) + 2(j-1) + 2.$$

Hence

$$|J_2| \leq C \left[\frac{1}{(j-1)j(k-j)} + \frac{1}{j(k-j-1)(k-j)} + \frac{1}{(j-1)j(k-j-1)(k-j)} \right] \leq C \left[\frac{1}{(j-1)^2(k-j-1)} + \frac{1}{(j-1)(k-j-1)^2} + \frac{1}{(j-1)^2(k-j-1)^2} \right]. \quad (3.9)$$

From (3.7)–(3.9) we obtain for $2 \leq j \leq k-2 \leq n-2$,

$$|a_{jk}| \leq \frac{C\epsilon}{(j-1)(k-j-1)} + C(a)\left[\frac{1}{(j-1)^3(k-j-1)} + \frac{1}{(j-1)(k-j-1)^3}\right] + C\left[\frac{1}{(j-1)^2(k-j-1)} + \frac{1}{(j-1)(k-j-1)^2} + \frac{1}{(j-1)^2(k-j-1)^2}\right]. \quad (3.10)$$

Now from [1],

$$\sum_{2\leq j\leq k-2\leq n-2} \frac{1}{(j-1)(k-j-1)} \sim (\log n)^2, \quad (3.11)$$

while

$$\sum_{j=2}^{n-2} \sum_{k=j+2}^{n} \frac{1}{(j-1)^2(k-j-1)} \leq \sum_{j=2}^{n-2} \frac{1}{(j-1)^2} \cdot \sum_{m=1}^{n} \frac{1}{m} = 0(\log n), \quad (3.12)$$

$$\sum_{j=2}^{n-2} \sum_{k=j+2}^{n} \frac{1}{(j-1)(k-j-1)^2} \leq \sum_{j=2}^{n-2} \frac{1}{j-1} \sum_{m=1}^{\infty} \frac{1}{m^2} = 0(\log n). \quad (3.13)$$

From (3.4)–(3.6) and (3.10)–(3.13) it follows that

$$\limsup_{n\to\infty} (\log n)^{-2} \left| \sum_{j,k=1}^{n} a_{jk} \right| \leq C\epsilon,$$

and since ϵ is arbitrary this proves (3.1) and Lemma 2.

Turning to the proof of Lemma 1, let

$$K_n = \frac{1}{\log n}\left[\int_0^n f(V(t))\,dt - \sum_{j=1}^{n} f(V(j))\right] = \frac{1}{\log n} \sum_{j=1}^{n} \int_{j-1}^{j} \left[f(V(t)) - f(V(j))\right] dt.$$

$$\text{Then } EK_n^2 = (\log n)^{-2} \cdot \sum_{j,k=1}^{n} b_{jk},$$

where

$$b_{jk} = \iint_{R_{jk}} E[f(V(t)) - f(V(j))][f(V(u)) - f(V(k))]\,dtdu$$

$$= \iint_{R_{jk}} [\{\theta(t,u) - \theta(t,k)\} - \{\theta(j,u) - \theta(j,k)\}]\,dtdu$$

and where we now write

$$\theta(t, u) = Ef(V(t))f(V(u))$$

$$= \frac{1}{4\pi^2 t(u-t)} \int \cdots \int f(x, y)f(\zeta, \eta)e^{-Q(t, u)}\, dx \ldots d\eta$$

for $0 < t < u$. For $2 \leq j \leq k-2 \leq n-2$,

$$\theta(t, u) - \theta(t, k) = \frac{1}{4\pi^2 t(u-t)} \int \cdots \int f(x, y)f(\zeta, \eta)[e^{-Q(t, u)} -$$

$$e^{-Q(t, k)}]\, dx \ldots d\eta + \frac{1}{4\pi^2 t}\left[\frac{1}{u-t} - \frac{1}{k-t}\right] \int \cdots \int f(x, y)f \times$$

$$(\zeta, \eta)e^{-Q(t, k)}\, dx \ldots d\eta = J_1 + J_2.$$

In R_{jk},

$$Q(t, u) - Q(t, k) = \frac{(x-\zeta)^2 + (y-\eta)^2}{2}\left(\frac{1}{u-t} - \frac{1}{k-t}\right) =$$

$$\frac{(x-\zeta)^2 + (y-\eta)^2}{2} \cdot \frac{k-u}{(u-t)(k-t)} \geq 0.$$

Hence $\left|e^{-Q(t, u)} - e^{-Q(t, k)}\right| = e^{-Q(t, k)}\left|e^{-\{Q(t, u) - Q(t, k)\}} - 1\right| \leq 1 -$

$$e^{-\{Q(t, u) - Q(t, k)\}} \leq Q(t, u) - Q(t, k) \leq \frac{(x-\zeta)^2 + (y-\eta)^2}{2} \cdot$$

$$\frac{k-u}{(u-t)(k-t)}.$$

Thus, writing

$$J_1 = \frac{1}{4\pi^2 t(u-t)} \int \cdots \int_{R_a'} + \frac{1}{4\pi^2 t(u-t)} \int \cdots \int_{R_a} = J_1' + J_1'',$$

we have $J_1' \leq \dfrac{C\epsilon}{(j-1)(k-j-1)}$,

$$J_1'' \leq \frac{F^2}{8\pi^2 t(u-t)} \int_{R_a} \cdots \int [(x-\zeta)^2 + (y-\eta)^2] \times$$

$$dx \ldots d\eta \cdot \frac{k-u}{(u-t)(k-t)} \leq \frac{C(a)}{(j-1)(k-j-1)^3},$$

$$|J_2| \leq \frac{C}{t} \cdot \frac{k-u}{(u-t)(k-t)} \leq \frac{C}{(j-1)(k-j-1)^2}.$$

Hence we have

$$|\theta(t, u) - \theta(t, k)| \leq \frac{C\epsilon}{(j-1)(k-j-1)} + \frac{C(a)}{(j-1)(k-j-1)^3} + \frac{C}{(j-1)(k-j-1)^2}, \quad (3.14)$$

and similarly it can be shown that the right-hand side of (3.14) is an upper bound for $|\theta(j, u) - \theta(j, k)|$. Hence, as in the proof of Lemma 2,

$$\lim_{n\to\infty} (\log n)^{-2} \sum_{2\leq j\leq k-2\leq n-2} b_{jk} = 0,$$

and the other sums occurring in EK_n^2 can be proved to be $o\{(\log n)^2\}$ as before, completing the proof of Lemma 1.

4. *The One-Dimensional Case.*—Let $f(x)$ and $g(x)$ be real valued functions which are bounded and summable in the line $-\infty < x < \infty$, and set $\bar{f} = \int f(x)dx$, $\bar{g} = \int g(x)dx$.

THEOREM 3. *If $\bar{f} \neq 0$ then for every u,*

$$\lim_{T\to\infty} \Pr\left[\frac{1}{\bar{f}\sqrt{T}} \int_0^T f(X(t))\, dt \leq u\right] = H(u),$$

where

$$H(u) = \begin{cases} \sqrt{\frac{2}{\pi}} \int_0^u e^{-y^2/2}\, dy & \text{for } u \geq 0, \\ 0 & \text{for } u < 0. \end{cases}$$

THEOREM 4. *If $\bar{g} \neq 0$ then*

$$\lim_{T\to\infty} \frac{\int_0^T f(X(t))\, dt}{\int_0^T g(X(t))\, dt} = \frac{\bar{f}}{\bar{g}} \text{ in probability.}$$

* John Simon Guggenheim Memorial Fellow.

OMNIBUS CHECKING OF THE 61-PLACE TABLE OF DENARY LOGARITHMS COMPILED BY PETERS AND STEIN, BY CALLET, AND BY PARKHURST

BY HORACE S. UHLER

YALE UNIVERSITY

Communicated by J. B. Whitehead, March 30, 1953

The first reference is: *Zehnstellige Logarithmentafel. Erster Band. Herausgegeben von Reichsamt für Landesaufnahme unter wissenschaftlicher Leitung von Prof. Dr. J. Peters.* Berlin 1922. Table 14b, pages 156–162 of the appendix. The original source for Table 14b is acknowledged on page xix by the statement that "···this table contains the 61-place common

logarithms of all the integers from 1 to 100 and above 100 of the primes alone up to 1097 according to the data of Callet (*Tables portatives de logarithmes, Paris 1795, an III*), $\cdots$."

The word "omnibus" in the title is intended to indicate that groups of one hundred or more logarithms were added and the sum compared with the logarithms of certain factorials conformably to the obvious relation $\log\{(n)(n-1)(n-2)\cdots(k+2)(k+1)\} = \sum_{k=1}^{n} \log l = a_1 \log 2 + a_2 \log 3 + \cdots + a_{l-1} \log l = \log(n!/k!) = \log(n!) - \log(k!)$, which may be conveniently solved for $\log(n!)$. The logarithms of the factorials were obtained from Stirling's asymptotic series $\log \Gamma(x) = (\log 2\pi)/2 + (x - {}^1/_2)\log x - x + \sum_{m=1} (c_m/x^{2m-1}) + R$ where $c_m \equiv (-1)^{m-1} B_m/[(2m-1)(2m)]$ and $\Gamma(x+1) = x!$ for x integral. It is here assumed that, in a carefully computed table, the probability of false digits cancelling in the sum of the logarithms is negligible. Emphasis should be placed on the fact that this investigation pertains primarily to internal errors and not to terminal figures. From 101 to 1097 the work was complicated by the absence from Table 14b of all composite numbers since the ordinary factorials contain consecutive numbers regardless of their prime or composite character. Again since the logarithms under test are on the base 10 while those computed by the author are for the most part on the base e the arithmetical labor was increased by the necessity of multiplying by the modulus $\log_{10} e$. The numerical data next presented in detail should clarify any inadequacy in the preceding outline of procedure.

We shall now consider only one typical case, namely, the calculations involved in the discovery of the error in log 839. For convenience in applying Stirling's series the boundary numbers 800 and 900 were chosen for one section of Table 14b. All of the natural numbers beginning with 801 and ending with 900 comprise the prime factors in the following product $2^{99} \cdot 3^{52} \cdot 5^{25} \cdot 7^{16} \cdot 11^{10} \cdot 13^{9} \cdot 17^{6} \cdot 19^{5} \cdot 23^{5} \cdot 29^{5} \cdot 31^{4} \cdot 37^{3} \cdot 41^{2} \cdot 43^{2} \cdot 47^{2} \cdot 53 \cdot 59^{2} \cdot 61 \cdot 67^{2} \cdot 71 \cdot 73^{2} \cdot 79 \cdot 83 \cdot 89^{2} \cdot 97 \cdot 101 \cdot 103 \cdot 107 \cdot 109 \cdot 127 \cdot 137 \cdot 139 \cdot 149 \cdot 163 \cdot 167 \cdot 173 \cdot 179 \cdot 211 \cdot 223 \cdot 269 \cdot 271 \cdot 277 \cdot 281 \cdot 283 \cdot 293 \cdot 401 \cdot 409 \cdot 419 \cdot 421 \cdot 431 \cdot 433 \cdot 439 \cdot 443 \cdot 449$. Curiosity calls attention to the fact that the following 33 primes in the interval did not occur, namely, 113, 131, 151, 157, 181, 191, 193, 197, 199, 227, 229, 233, 239, 241, 251, 257, 263, and all three digit primes whose common leading figure was 3. Primes greater than 449, such as 457, 463, $\cdots$, are not admissible because each exceeds one-half of 900. By having Table 14b photostated and then slicing the negatives into narrow strips, accuracy was increased and time was saved either by juxtaposition or by machine adding of selected strips arranged under and flattened by a sheet of plate glass. In calculating log(900!) from Table 14b it was necessary to borrow from this table the logarithms

of numbers greater than 127 because the author's own tables did not include the desired data. This procedure was quite justifiable because in the earlier part of the present investigation it was shown that only one interior error existed in Table 14b for entries less than log 457. This error falls in the 58th decimal place of the mantissa of log 227 where the 12th pentad is printed as "49465" instead of the correct group 49565. Stirling's series gave log(900!) = 2269. 82947 61838 14835 45315 15825 70310 51541 48578 90591 62162 01337 59483 25989···. The value of this logarithm as obtained from Table 14b agreed perfectly with the number just recorded through the 57th decimal place. But the digits beginning with the 56th place were 59582 7. Hence the excess 00099 4 of the tabular sum over the true value indicated an error of one unit in the 58th place in some one of the printed logarithms of the 15 primes 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, and 887. The author's zero experience with games of chance caused him to calculate at random the logarithms of 809, 827, 853, 857, 859, and 863 before the error of +1 was run down in log 839. The same innocence led to the scrambled calculation of the logarithms of 1013, 1021, 1031, 1033, 1049, 1063, 1093, and 1097. Two errors were unveiled in log 1097.

To recapitulate, all of the errors in Table 14b known to the author are collected in table 1.

TABLE 1

LOG OF	DEC. PLACE	*for*	*read*
127	61	2	3
227	58	49465	49565
293	61	3	4
839	58	53974	53874
1009	59	38238	38228
1097	58	00941	00941
1097	61	8	7

The author does not claim priority for the discovery of all of these errors but he does desire to stress the fact that the procedure explained in the earlier paragraphs definitively proves that no other non-cancelling interior errors exist in Table 14b.

The extensions of the natural and denary logarithms beyond the 48th and 61st decimal places, respectively, as found in the present work, deserve recording. For the former see Tafel 13. In table 2 the last digits have been conventionally "rounded off."

Juxtaposition of the sliced photostatic negatives of the 61-place tables of Callet against Table 14b of Peters and Stein showed that the latter had copied the earlier table perfectly, that is, including errors, last digits, and every detail. This comparison indicates that Peters and Stein did not fulfill the promises made or attain the standard of accuracy proposed in the

following translation from page III of the appendix. "While examining the pertinent literature it appeared to us that the published data were not always correct and they afforded no guarantee that the accuracy of one half a unit in the last decimal figure as aspired to by us was actually attained. Accordingly as a matter of principle we have calculated anew all of the numbers presented in the following pages. Even in the few cases in which the recalculation of the published data as far as the last decimal would have required an inordinate amount of time so that it has been omitted, such as for example the constant π with 707 places, nevertheless the values have been recalculated at least in their initial figures (some 50 to 60 decimals) and the remaining decimals have been transcribed from the respective authors after the application of appropriate checks." The following verbatim quotation seems to be self-incriminating. Page V, Section 3.

TABLE 2

N	LN N 46TH PL. ONWARD	LOG N 56TH PL. ONWARD
293	89170,89475,47248,42893,11916,277	38729,39794,53776,4460
809	48300,59089,16663,95699,34535,089	02266,18182,60534,4009
827	29126,45785,33904,60968,68762,048	57998,22870,51816,6934
839	54281,75400,47382,99126,14778,660	53874,12199,35504,4051
853	02716,27159,24195,03767,99674,090	38386,83091,45246,2237
857	79717,94675,50493,18461,50815,105	56694,64756,51171,1988
859	25879,46508,86537,58084,72940,592	60098,73672,29372,7937
863	46602,38093,94780,30999,70284,132	02396,17110,47809,3550
1013	71829,94572,08322,66959,06295,926	25895,37720,01904,5775
1021	07072,05161,50322,14128,60597,592	51893,63435,12531,1482
1031	80558,68462,65744,35802,98414,342	30173,04759,00572,0830
1033	79426,18876,27458,07365,89581,391	27216,15291,12745,0294
1049	23900,76540,21424,45926,10403,259	21217,27842,03186,9215
1063	39965,32348,96201,63980,68427,869	29966,18299,06500,3304
1093	03983,28952,06228,18694,22395,475	01056,14226,25516,3272
1097	75652,65165,11623,42347,77247,408	00941,74706,58168,3747

"Mit Rücksicht auf die 61-stelligen Logarithmen bei Callet (hier Tafel 14b) haben wir sie 61-stellig gegeben und ihre Richtigkeit in gleicher Weise wie oben (s. u. Ziffer I) durch wirksame Kontrollen (auch hinsichtlich der Endziffern) gesichert." The book by François Callet had the title *Tables Portatives de Logarithmes: Paris 1795* (Tirage 1846).

The author takes pleasure in gratefully acknowledging the temporary loan, as effected by his colleague Professor R. C. Archibald, of the volume containing the tables by Parkhurst. This booklet is the property of the library of Brown University. Comparison of the 61-place table with that of Peters and Stein showed the identical errors recorded above. Parkhurst doubtless copied Callet's data for he wrote in the table of contents "XVIII Logarithms to 61 decimals, from Cala,···, p. 78." Much credit redounds to Parkhurst for having devised home-made euphonic English and for

assembling some of his numerical data in the most inconvenient and confusing manner imaginable!

For additional information the reader may consult the following references.

[1] Archibald, R. C., and Uhler, H. S. "Errors in the Tables of Peters and Stein in the Anhang to J. T. Peters Zehnstellige Logarithmentafeln, Vol. 1," *Math. Tables and Aids to Computation*, **1**, No. 2, pp. 57–59, April, 1943.

[2] Uhler, H. S., "Errors in Parkhurst's 100-Place Tables of Logarithms," *Math. Tables and Other Aids to Computation*, **1**, No. 4, pp. 121, 122, October, 1943.

[3] Archibald, R. C., "New Information Concerning Isaac Wolfram's Life and Calculations," *Math. Tables and Other Aids to Computation*, **4**, No. 32, pp. 185–209, October, 1950.

*SIGNIFICANCE LEVELS FOR A SKEW DISTRIBUTION**

BY EDWIN B. WILSON

OFFICE OF NAVAL RESEARCH, BOSTON

Communicated March 30, 1953

1. The Symmetric Distribution.—The first time one meets the problem of significance at about the 0.05 level is when one first writes $x \pm 2\sigma$, or, if dealing with a rate construed as a probability, $p \pm 2(pq/n)^{1/2}$. The implication is that a statistical fluctuation from the value x or p will occur to an absolute amount of twice the standard deviation of x or p only about once in 20 times, or once in 40 times on either side. In a strictly normal distribution[1] the proper multiplier for the level 0.05 is 1.96, not 2. Various refinements up to Student's t-test are familiar.

In the case of comparing two rates, p_1 and p_2, in samples of n_1 and n_2, the significance test is often put in the form

$$\chi^2 = \frac{(p_1 - p_2)^2}{p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} > 3.84 \quad \text{or} \quad \frac{|p_1 - p_2|}{\sqrt{p_0 q_0}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > 1.96.$$

The χ^2 distribution extends from 0 to ∞ and, for one degree of freedom, is the positive half of the normal distribution doubled, viz.,

$$f(z) = \sqrt{\frac{2}{\pi}}\, e^{-z^2/2}, \qquad 0 \leqq z = \sqrt{\chi^2} < \infty.$$

Of course, if $f(z)$ is the distribution of z from $-\infty$ to ∞ about $z = 0$ and if $f(z) = f(-z)$, the distribution of the absolute value $|z|$ must be $2f(|z|)$ and run from 0 to ∞. The level of significance of a value z is refined as

the area under the distribution function $f(z)$ more remote from $z = 0$ than $\pm z$, viz.:,[2]

$$\int_{-\infty}^{-z} f(z)\,dz + \int_{z}^{\infty} f(z)\,dz = 2\int_{z}^{\infty} f(z)\,dz, \qquad z > 0.$$

It can have any value from 0, when the stated value z is at the ends of the distribution, to 1, when it is at the center. All this is quite familiar: there seems to be universal agreement both in theory and in practice in regard to it, even though the words used may differ from one to another.

2. The Target Problem.—If one considers the two-dimensional χ^2 distribution in the form

$$P(\chi^2 = r^2) = \int_{r}^{\infty} e^{-r^2/2} r\,dr = e^{-r^2/2}, \qquad (1)$$

which arises in the Bravais target problem for a circular pattern, the mode of the distribution in r is at $r = 1$, the median at $r = (2 \log_e 2)^{1/2} = 1.177$, the mean at $(\pi/2)^{1/2} = 1.253$, and the root mean square at $r = \sqrt{2} = 1.414$—this last being the radius of gyration of the hits. If one wishes to estimate the bad shots in terms of the probability of a deviation so great from the bull's-eye $r = 0$ and at the usual significance level 0.05, one has only to solve $e^{-r^2/2} = 0.05$ for r to find $r^2 = 5.991$ and $r = 2.447$.

Although in the ballistic problem one would probably be interested in deviations from $r = 0$, one might have precisely the same distribution for r in a problem in which the interest was in deviations from some specified value of r such as $r = 1.414$. Indeed in the kinetic theory of gases it is the kinetic energy of the molecules which is of greatest interest and the distribution according to velocity is the three-dimensional counterpart of (1). And one might very well be asking about the molecules, say, the 5% of molecules, whose velocity departed most from the root mean square velocity which determines the mean kinetic energy, and thus the temperature. These aberrant molecules would be of two sorts, those with velocity well above and those with velocity well below the particular velocity specified, whether it be the root mean square, mean, median, mode, or any other.

To have a name for it, I shall call the specified value the *target*[3] *value;* it may be any value within the range of possible values. If T be the target value and $r_0 < T < r_1$ the condition for significance at the level L is by definition

$$\int_{0}^{r_0} e^{-r^2/2} r\,dr + \int_{r_1}^{\infty} e^{-r^2/2} r\,dr = L. \qquad (2)$$

3. Discussion of the Definition.—If $L = 0$, the lower limit r_0 must be the lowest limit of the range of the distribution and the upper limit the uppermost; if $L = 1$, $r_0 = r_1 = T$. If the lowest limit of the distribution is finite and the target value coincides therewith, $r_0 = 0$ and r_1 is determined by (2); and similarly if the upper limit of the distribution were finite and T

coincided with it, r_1 would equal T and r_0 would be fixed. In most other cases, r_0 and r_1 would either be indeterminate or must be fixed by some special convention to be adopted only after critical discussion of what is appropriate to the problem in hand.

(*a*) For instance, if one is concerned with the actual absolute value of the difference $|r - T| = h$, one should take $r_0 = T - h$ and $r_1 T + h$ for all values of h which do not carry beyond the range of the frequency function with the understanding that if either r_0 or r_1 reaches one end of the range before the other, that other shall be let run on to its end of the range. This procedure seems so natural for the discussion of the chance variation of the number of deaths in a population subject to a low death rate that it is almost universally adopted.

(*b*) Secondly, it is often the practice to split the level L into two equal parts and apply each half to each tail of the frequency function. If this procedure is to be followed right through for all levels from 0 to 1, it is necessary that the target value be the median; otherwise one would apply it only until either r_0 or r_1 closed down on the target and would then put all the rest of the level into the other tail.

(*c*) In the third place, one might use the condition $f(r_0) = f(r_1)$ in the usual unimodal frequency function with tails running to 0 at the ends of the range, so that the definition of significance would cover all cases less frequent than $f(r_0) = f(r_1)$. Here, if consistently carried through to $L = 1$, the target value must be the mode, or some adjustment must be made as in the two previous instances. An alternative statement of the condition would be to require the range $r_1 - r_0$ to be a minimum.

(*d*) In the fourth place, one might divide L in proportion to the areas under the frequency function from T to the two ends of the range. Thus in the case of (1), if one were discussing significant deviations from $r = T = 1.414$, the radius of gyration, which corresponds to the molecular velocity of the mean kinetic energy in the three-dimensional case of a gas, one would first find that the probability of a shot falling outside $r = 1.414$ is $1/e = 0.368$ and would then determine r_0 and r_1 by the equations

$$\int_0^{r_0} e^{-r^2/2} r\, dr = 0.632\, L, \qquad \int_{r_1}^{\infty} e^{-r^2/2} r\, dr = 0.368\, L,$$

so that $r_0 = 0.252$ and $r_1 = 2.83$, if $L = 0.05$.

All these four possibilities coalesce into a single one if the distribution is symmetrical; none seems to have been generally adopted; and there are still other possibilities. Of the four, the second and fourth are invariant under a change of variable, which is distinctly in their favor if and only if the variable is merely ordinal without any cardinal significance; and the fourth alone can be applied in general without modification all the way from $L = 0$ to $L = 1$. There is another property of the fourth which deserves noting, namely: the (relative) probability that r be less than r_0 if

less than T and the (relative) probability that r be greater than r_1 if greater that T, and the probability that r lie outside the interval $r_0 < r < r_1$ are all the same.

Whether this property is important may be judged by asking what would be the appropriate answer to the question: How to define significance level in case the only interest is in a part of the possible range of the variable? To be concrete, what should be the 0.05 significance level of wildness of shots outside the ring $r = 1.414$ around the bull's-eye? If we reply by taking an area 0.05 off the infinite end of the whole distribution we get $r = 2.45$; and we must get the same result if we had specified any other ring $r = T \leqq 2.45$ instead of $r = 1.414$, so that from our specifications the "outside the ring $r = 1.414$ around the bull's-eye" might better have been omitted. The question as actually put might better be answered on the basis of relative probabilities by $r = 2.83$.

4. *Variance Analysis.*—The fundamental frequency function is

$$\frac{n_1^{n_1/2} n_2^{n_2/2}}{B(n_1/2, n_2/2)} \frac{F^{n_1/2-1}}{(n_1F + n_2)^{(n_1+n_2)/2}}$$

or

$$\frac{n_1^{n_1/2} n_2^{n_2/2}}{B(n_1/2, n_2/2)} \frac{2e^{n_2 z}}{(n_1e^{2z} + n_2)^{(n_1+n_2)/2}}$$

where F is the ratio of two variances s_1^2/s_2^2 of n_1 and n_2 degrees of freedom, and z is half its natural logarithm. The distribution of z has a mode at $z = 0$, which corresponds to equality of the two observationally determined variances; but the z-distribution is skew[4] unless $n_1 = n_2$. The question is: Is the ratio F so large or so small that its observation implies that the assumption that the two samples came from universes with equal variances is untenable at a specified level L of significance?

The target value of F is not 1, nor of z the correlative value 0; it is taken to be the median, which is that value which F (or z) is as likely as not to exceed if the samples come from universes of equal variances; and the significance level L is divided equally between the two tails as in the second possibility listed above. For example, if one takes $L = 0.10$ and $n_1 = 1$, $n_2 = 2$, one finds $z_0 = -2.6479$ and $z_1 = 1.4592$ with corresponding values for F, as recorded in the standard published tables.[5] The median is -0.2028 for z and $^2/_3$ for F.

The t-test for the mean, being symmetrical, can be stated as a test for $F = t^2$, varying from 0 to ∞, and when so stated the frequency function becomes identical with that for the ratio of two variances s_1^2 with one degree of freedom and s_2^2 with n; but this does not mean that the t-test thus formulated is the same as a variance test,[6] because for the former the target value is $T = 0$ whereas for the latter it is the median F.

It is well known that the last row in the table for z, when $n_2 = \infty$, gives the value $z = {}^1/_2 \log (\chi^2/n_1)$ for 0.05. This is as it should be because with $n_2 = \infty$ the value of s_2^2 must be σ_2^2, the variance of the universe, and the ratio F then becomes simply χ^2/n_1. The target value from which χ^2 is normally measured is 0 and that of z is correspondingly $-\infty$. This in itself is sufficient to show that the 0.05 table for z does not give the values of z at the 0.05 level of significance in variance analysis, for the target value of z is then not $-\infty$;[7] the 0.05 table for z gives one measure of the 0.10 significance level. This, I believe, many find confusing.

5. A Bit of History.—The matter of the z distribution first came to the attention of a wide population of statisticians when the first edition of R. A. Fisher's *Statistical Methods for Research Workers* appeared in 1925. It was in Art. 41, pp. 192–193, that he wrote:

> These tests are all reducible to the single problem of testing whether one estimate of variance derived from n_1 degrees of freedom is significantly greater than a second such estimate derived from n_2 degrees of freedom. This problem is reduced to its simplest form by calculating z equal to half the difference of the *natural* logarithms of the estimates of the variance, or to the difference of the logarithms of the corresponding standard deviations. Then if P is the probability of exceeding this value by chance, it is possible to calculate the value of z corresponding to different values of P, n_1, and n_2.
>
> . . . When both n_1 and n_2 are large, and also for moderate values when they are equal or nearly equal, the distribution of z is sufficiently nearly normal for effective use to be made of its standard deviation, which may be written $\sqrt{\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$.

If we take $n_2 = n_1 = 24$, the standard deviation of z is 0.2041 and 1.96 times this is 0.4001, which would correspond to the 0.05 significance level in a normal curve; but the entry in the table is 0.3425, and this does in fact correspond to the 0.10 level in the normal curve.[8] The fact that Fisher's rule for the standard deviation in his second paragraph does not give the values in the table prepared in accord with his first, has perplexed a number of students, and has apparently been overlooked by many others.[9]

One may use the ratio $R_1 = z/\sigma_z$ as a normal variate to estimate significance, as Fisher says. In the old days before Fisher invented the rigorous theory of variance analysis the manner of comparing s_1 and s_2 was by a difference as was, and still is, the case for two means; this would lead to the critical ratio

$$R_2 = \frac{s_1 - s_2}{\sigma\sqrt{\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{with} \qquad \sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}.$$

Like the ratio R_1, it would apply only when n_1 and n_2 were equal and moderate or both fairly large. If recourse be had to the result that

$$\left[\left(\frac{\chi^2}{n}\right)^{1/3} - \left(1 - \frac{2}{9n}\right)\right] \div \sqrt{\frac{2}{9n}}$$

is approximately a normal variate of unit variance over the range from -0.8 to 2.4, even when n is small,[10] one might set up a third ratio with some hope that it would give a good approximation over a considerable range even for tolerably small numbers n and for numbers which were different, viz.:

$$R_3 = \left[\frac{s_1^{2/3} - s_2^{2/3}}{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\right)^{1/3}} + \frac{2}{9}\left(\frac{1}{n_1} - \frac{1}{n_2}\right)\right] \div \sqrt{\frac{2}{9}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

To see how these three R's compare at $P = 0.10$ when $n_1 = n_2$, one need only substitute for z in Fisher's table VI (labeled $P = 0.05$) or Kendall's table 4, *op. cit.*, p. 442, or any equivalent.[11] One finds:

	$n = 1$	$n = 2$	$n = 3$	$n = 6$	$n = 12$	$n = 24$
R_1	2.54	2.08	1.93	1.78	1.71	1.678
R_2	1.30	1.50	1.56	1.61	1.63	1.638
R_3	1.54	1.643	1.657	1.659	1.653	1.649

The proper value for R is 1.645. The values of R_1 are too large, those of R_2 too small, and those of R_3 sometimes too large and sometimes too small. Whether one should compare the deviations or turn them into the corresponding probabilities, and how to compare the probabilities, are questions that would have to be discussed, and there might well be different opinions as to the rigorously correct answers; but the display seems to indicate fairly clearly that on the whole R_1 is the worst and R_3 is the best of the tests; indeed for equal n's, R_3 gives satisfactory results from $n = 2$ up, and the old fashioned R_2 from $n = 12$ up—more satisfactory than R_1 gives for $n = 24$.

If next one applies R_3 to $n_1 = 2$ and a variety of n_2's, one has:

	$n_2 = 3$	$n_2 = 4$	$n_2 = 5$	$n_2 = 8$	$n_2 = \infty$
R	1.676, -1.623	1.682, -1.607	1.682, -1.580	1.678, -1.578	1.659, -1.553

It will be seen that the positive limit increases a bit and then decreases but remains close to the proper value 1.645, whereas the negative limit falls off steadily to about -1.55, and the trial of other cases seems to confirm this behavior. If one goes over to $n_1 = 3$, R_3 for $n_2 = \infty$ is -1.604 and for $n_2 = 4$ it is -1.622. In the vicinity of this point of the normal curve a drop of 0.01 in the abscissa brings a change to $P = 0.12$, or about 20%. It looks safe to say that for n_1 and n_2 both greater than 3, R_3 will reproduce the table for $P = 0.10$ with an accuracy of within about 5%, i.e., P will lie between 0.095 and 0.105.

It should be noted that it is upon the negative side that the errors of R_2 are the more serious; this corresponds to the fact that the approximation to χ^2 is worse on the negative side, i.e., for values near 0. It is therefore to be expected that if one desired to work at the significance level $P = 0.05$ one would have to compute a table for that level up to the point where R_2 would give good enough values or one would have to find an improved approximation to χ^2 and base a test upon that; or one would have to make some adjustment in the expression for R_3 as given.[12]

Summary.—The various desires of statisticians treating a great variety of problems will probably result in a considerable variety of definitions for significance. So long as there remain differences of definition care should be exercised to make sure that the definition adopted in any particular problem is appropriate to that problem and is made clear.

It seems that there is involved in the concept of significance level something I have called a target value of zero significance from which departures are measured on some sort of scale; they may, according to the problem, be measured in either direction or in both.

In case there is no cardinal but only an ordinal meaning to deviations from the target value, the frequency function $f(z)$, with its integral distribution function I, where $dI = f(z)\,dz$, may be transformed to any variable. In particular if I is used as the variable, the new frequency function becomes identically 1 over the range from 0 to 1, and should the variable $II = 2(I - {}^1/_2)$ be introduced, the function would be constant at ${}^1/_2$ over the range -1 to $+1$. If the target value were the median, the frequency function would then be symmetric with respect to its median and the skewness of the original distribution may be regarded as accidental.[13] If the target value be not the median, the function $f(z)$ can still be transformed, at least conceptually, to the uniform value ${}^1/_2$ over the range -1 to $+1$, but as the target value will not be at the midpoint of the range, the original distribution $f(z)$ taken with its target value $z = T$, has to be considered as essentially asymmetric. This is the case which offers the most serious liability to difference of definition of significance level between different statisticians.

While it is my opinion that as yet no unitary definition or set of alternative definitions can be imposed on the profession, I would urge consideration of my fourth definition in §2 (d),[14] so that the significance assigned to II relative to the target II_0 be the same whether the interest is only in $II > II_0$ or $II < II_0$ or in all values of II; and I should think that such must be the case if whether II be greater or less than II_0 be the result of a convention applicable only after the calculations have been made, as if it should be agreed that the difference of two means should be so taken after their determination as to be positive.

Some of these considerations were applied to that type of variance

analysis in which the problem is the acceptance or rejection at a given significance level of the hypothesis that the samples came from universes with equal variances. A new approximate test therefor was suggested.

* This is the third note on points in elementary statistics which seem not beyond the stage of discussion; previous ones in these PROCEEDINGS, **37,** 696–704 (1951) and **38,** 899–905 (1952).

[1] It seems to be customary to use the term distribution in two senses: (1) for the frequency distribution $f(z)$, and (2) for the integral distribution function $I(z)$ thereto related by $dI = f\,dz$.

[2] More generally, the limits may be $\pm a$; one may still write the limits as $\pm\infty$ if one will adopt the convention that $f(z) \equiv 0$ outside its meaningful range.

[3] One should not confuse the target value with the bull's-eye.

[4] The ratio $f(z)/f(-z)$, $z > 0$, runs monotonically from 1 when z is 0 to 0 or ∞ when z is infinite according as $n_1 < n_2$ or $n_1 > n_2$.

[5] The tabulated values of z are for the 0.05 point of the (integral) distribution function on the hypothesis that the larger variance is in the numerator and thus those values are all positive; an interchange of numerator and denominator changes the sign of z. Thus $z_0 = -2.6479$ for $n_1 = 1$ and $n_2 = 2$ is found in the tables as $z = 2.6479$ for $n_1 = 2$ and $n_2 = 1$.

[6] It is instructive to deal with this in some detail. The usual form of the frequency function for the t-test is

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, \quad -\infty < t < \infty,$$

or

$$\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, \quad 0 < t < \infty.$$

In either case the target value is 0; in the former at the center of the range, in the latter at the extreme left. In this latter case one could introduce $e^z = t$ and write

$$\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{e^z}{(1+e^{2z}/n)^{(n+1)/2}}, \quad -\infty < z < \infty,$$

which is the form of the z distribution for $n_1 = 1$ with the proper factor of $n = n_2$; but the target value remains at the extreme left, $z = -\infty$. This z-distribution is itself symmetrical around $z = 0$ when and only when $n = 1$, in which case it becomes $(2/\pi)e^z/(1 + e^{2z})$. For the t-test it must be left in this form with $z = -\infty$ as target value; but for the variance test with its target value at $z = 0$, it may be doubled and restricted to $z > 0$. Thus for the variance test with $n_1 = n_2 = 1$, one has

$$\frac{4}{\pi}\frac{e^z}{1+e^{2z}}, \quad 0 < z < \infty, \qquad \text{and} \qquad \frac{4}{\pi}\left(\frac{\pi}{2} - \tan^{-1} e^z\right) = L.$$

If $L = 0.05$, the result is $z = 3.237$; if $L = 0.10$, it is 2.542 as entered in the tables for the 0.05 point.

[7] I am speaking only of that type of variance analysis with which I have been most concerned in various biological, medical, and social applications with which I have been busied, viz., where the two samples are on an equal footing and the problem requires the acceptance or rejection at the significance level L of the "null hypothesis" that in the universes from which the samples came the variances are equal. The derivation of the frequency distribution is found in M. G. Kendall's *Advanced Theory of Statistics*, Vol. I, p. 249 (unfortunately with some typographical errors in the original edition, but happily not serious ones). For the variance analysis of which I am talking the target value of χ^2/n_1 is not 0 but the median of χ^2/n_1 in a universe whose variance is σ_2^2.

[8] The extreme difference from $n_1 = n_2$ is when $n_1 = 1$, $n_2 = \infty$. For this the frequency function is $(2/\pi)^{1/2}e^{-u^2/2}$ with $u^2 = F = \chi^2$, if the variance with one degree of freedom be the numerator of F. The median is the "probable error" $u = 0.6745$ to which correspond targets $z = -0.394$ and $F = 4/9$. The integral distribution function has the value 0.683 when $z = 0$, so that better than $^2/_3$ of the values of s_1^2 are less than σ_2^2, and an observed value $s_1^2 = \sigma_2^2$ would indicate that the hypothesis $\sigma_1 = \sigma_2$ is on the border of rejection for $L = 0.634$; the same value of L would be indicated for $F = 0.167$ and $z = -0.894$ as for $F = 1$ and $z = 0$.

[9] To do adequate justice to Fisher in dealing with this matter I must call attention to the initial words of the quotation "These tests," for it is indeed possible that the tests to which he is there specifically referring are like the ordinary χ^2 test in that for them the target value is 0 and all the value of L should go into one tail of the distribution. And he may intend that when the standard deviation is used, it shall be considered to apply only on one side of $z = 0$, and to the amount 1.645σ instead of 1.96σ. Furthermore it may well be that the two problems, one asymmetric, of determining whether one estimate of variance is significantly greater than another, and the other symmetric, of examining a difference in terms of its standard deviation, are in fact different, though I find no explicit indication in these paragraphs that they are. I shall not, however, at this time go into this question nor into the more complicated matter of first and second kinds of error (a discussion of which can be found in Chap. VIII of *Techniques of Statistical Analysis* by C. Eisenhart, M. W. Hastay, and W. A. Wallis, McGraw-Hill, 1947). I am not trying to prove anybody wrong or right; there seems still to be too little agreement and too little clarity for that, and too much possibility that there is involved a variety of really different problems.

[10] Wilson, E. B., and Hilferty, M. M., these PROCEEDINGS, **17**, 684–688 (1931). M. G. Kendall did us the honor to give this approximation in his *Advanced Theory of Statistics*, Vol. I, §12.8, and to show how good it was at $P = 0.01, 0.05, 0.95, 0.99$ for $n = 20, 30, 40, 50$. We had no interest in its excellence except for small values of n. For $n = 1$ it is not good at $P = 0.8$, and although the least value χ^2 can have is 0, at which P should be 1.00, the approximation for $n = 1$ makes χ^2 vanish for the normal deviate -1.65, for which $P = 0.95$ instead of 1.00.

[11] In terms of F or z we may write

$$\frac{s_1^{2/3} - s_2^{2/3}}{\left(\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\right)^{1/3}} = \frac{2\sinh(z/3)}{\left(\cosh z + \dfrac{n_1 - n_2}{n_1 + n_2}\sinh z\right)^{1/3}} = \frac{F^{1/3} - 1}{\left(\dfrac{n_1 F + n_2}{n_1 + n_2}\right)^{1/3}},$$

the computation of which will be easier in the one or the other form according to what computational aids are available. The values tabulated were obtained with the use of Huntington's *Four Place Tables* and the final digits are not trustworthy.

[12] A table for $P = 0.05$, in our sense is given by W. J. Dixon and F. J. Massey, Jr., in

Introduction to Statistical Analysis, McGraw-Hill, New York, 1951, p. 310, with credit to E. S. Pearson, *Biometrika*, **33** (1943).

[13] The case $n_1 = 2$, $n_2 = n$ will not be symmetric in z unless $n = 2$, but it is immediately integrable in a very simple way:

$$H = \frac{1}{2} - \frac{1}{\left(1 + \frac{2F}{n}\right)^{n/2}}, \quad F = \frac{n}{2}\left[\left(\frac{2}{1-H}\right)^{2/n} - 1\right], \quad \text{median } F = \frac{n}{2}(2^{2/n} - 1).$$

The condition for significance at the level L is simply $|H| = 1 - L$. Clearly H will be positive or negative according as F is above or below its median.

[14] Assuming, of course, that we are dealing with variables essentially ordinal; that the cases in §2(a) and §2(c) have cardinal applications is apparent.

GENERALIZED RIEMANN SPACES AND GENERAL RELATIVITY

By Luther P. Eisenhart

Fine Hall, Princeton University

Communicated April 13, 1953

A space of coordinates x^i $(i = 1, \ldots, n)$ with which is associated a non-symmetric tensor g_{ij} whose determinant g is not zero, is termed a generalized Riemann space. By definition

$$\Delta_{ijk} = \frac{1}{2}(g_{ik,\,j} + g_{kj,\,i} - g_{ij,\,k}), \tag{1}$$

where throughout this paper a quantity followed by a comma and index indicates the derivative of the quantity with respect to the coordinate with that index; thus $g_{ik,\,j} = \partial g_{ik}/\partial x^j$.

Throughout this paper g_{ij} denotes the symmetric part of g_{ij}. It is assumed that the determinant of g_{ij} is not zero. Quantities g^{ik} are uniquely defined by

$$g_{ij}g^{ik} = \delta^k_j,$$

where the summation convention of a repeated index, as i in this case, is applied here and throughout the paper, and δ^k_j are Kronecker deltas. By definition

$$\Delta^h_{ij} = g^{hk}\Delta_{ijk}, \qquad \Delta_{ijk} = g_{hk}\Delta^h_{ij}. \tag{2}$$

It can be shown[1] that these quantities in two coordinate systems x^i and x'^α are in the relation

$$\frac{\partial^2 x^h}{\partial x'^\alpha \partial x'^\beta} + \Delta^h_{ij}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} = \Delta'^\gamma_{\alpha\beta}\frac{\partial x^h}{\partial x'^\gamma}. \tag{3}$$

From (1) one has

$$\begin{aligned} \Delta_{\underline{ijk}} &\equiv \frac{1}{2}(\Delta_{ijk} + \Delta_{jik}) = \frac{1}{2}(g_{\underline{ik},j} + g_{\underline{kj},i} - g_{\underline{ij},k}), \\ \Delta_{\underline{ijk}} &\equiv \frac{1}{2}(\Delta_{ijk} - \Delta_{jik}) = \frac{1}{2}(g_{\underline{ik}} + g_{\underline{kj}} + g_{\underline{ik}}), \end{aligned} \tag{4}$$

where $g_{\underline{ij}}$ denotes the skew-symmetric part of g_{ij}. Denoting by $\Delta^h_{\underline{ij}}$ and $\Delta^h_{\underline{ij}}$ the symmetric and skew-symmetric parts respectively of Δ^h_{ij}, we have

$$\Delta^h_{\underline{ij}} = \{^h_{ij}\}, \tag{5}$$

where $\{^h_{ij}\}$ is the Christoffel symbol of the second kind in the quantities $g_{\underline{ij}}$,[2] and

$$g_{\underline{hk}}\Delta^h_{\underline{ij}} = g_{\underline{hl}}\Delta^h_{\underline{jk}} = g_{\underline{hj}}\Delta^h_{\underline{ki}} = \Delta_{\underline{ijk}}. \tag{6}$$

From (2), (4) and (6), one has

$$\Delta_i \equiv \Delta^j_{\underline{ij}} = g^{\underline{jk}}\Delta_{\underline{ijk}} = g^{\underline{jk}}\Delta_{\underline{jki}} = 0. \tag{7}$$

If in equation (3) the indices α and β are interchanged and also the dummy indices i and j, one obtains

$$\frac{\partial^2 x^h}{\partial x'^\beta \partial x'^\alpha} + \Delta^h_{\underline{ji}}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} = \Delta'^\gamma_{\underline{\beta\alpha}}\frac{\partial x^h}{\partial x'^\gamma}. \tag{8}$$

When this equation is subtracted from equation (3) it is found that $\Delta^h_{\underline{ij}}$ is a tensor.

Another set of coefficients of affine connection Γ^h_{ij} are related in two coordinate systems x^i and x'^α by equations of the form (3), that is

$$\frac{\partial^2 x^h}{\partial x'^\alpha \partial x'^\beta} + \Gamma^h_{ij}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} = \Gamma'^\gamma_{\alpha\beta}\frac{\partial x^h}{\partial x'^\gamma}. \tag{9}$$

Subtracting equation (3) from this equation, one has

$$(\Gamma^h_{ij} - \Delta^h_{ij})\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} = (\Gamma'^\gamma_{\alpha\beta} - \Delta'^\gamma_{\alpha\beta})\frac{\partial x^h}{\partial x'^\gamma},$$

from which it follows that, if we put

$$\Gamma^h_{ij} = \Delta^h_{ij} + a^h_{ij}, \tag{10}$$

the quantities a^h_{ij} are components of a tensor.

It is now required that the tensor a^h_{ij} be such that

$$g_{ij,k} = g_{hj}\Gamma^h_{ik} + g_{ih}\Gamma^h_{kj}. \tag{11}$$

From this equation it follows that

$$g_{ij,k} - g_{hj}\Delta^h_{ik} - g_{ih}\Delta^h_{jk} = g_{hj}a^h_{ik} + g_{ih}a^h_{jk} + g_{hj}(\Delta^h_{ik} + a^h_{ik}) + g_{hi}(\Delta^h_{jk} + a^h_{jk}),$$

where a^h_{ik} and a^h_{ik} are the symmetric and skew-symmetric parts respectively of a^h_{ik}, similarly for a a^h_{jk}, and in what follows. In consequence of (5) the left-hand member of this equation is identically zero. Consequently the tensor a^h_{ij} must satisfy conditions of the form

$$g_{hj}a^h_{ik} + g_{hi}a^h_{jk} + g_{hj}(\Delta^h_{ik} + a^h_{ik}) + g_{hi}(\Delta^h_{jk} + a^h_{jk}) = 0.$$

Permuting i, j, k cyclically twice, one obtains respectively

$$g_{hk}a^h_{ji} + g_{hj}a^h_{ki} + g_{hk}(\Delta^h_{ji} + a^h_{ji}) + g_{hj}(\Delta^h_{ki} + a^h_{ki}) = 0,$$

$$g_{hi}a^h_{jk} + g_{hk}a^h_{ij} + g_{hi}(\Delta^h_{kj} + a^h_{kj}) + g_{hk}(\Delta^h_{ij} + a^h_{ij}) = 0.$$

One-half the sum of these three equations is

$$g_{hj}a^h_{jk} + g_{hj}a^h_{ki} + g_{hk}a^h_{ij} = 0. \tag{12}$$

By means of this result the first of the above equations may be replaced by

$$g_{hk}a^h_{ij} = g_{hj}(\Delta^h_{ik} + a^h_{ik}) + g_{hi}(\Delta^h_{jk} + a^h_{jk}). \tag{13}$$

When the expressions of the form (11) with the values (10) of the Γ's are substituted in equation (1), the result is reducible to

$$g_{hj}(\Delta^h_{ki} + a^h_{ki}) + g_{ih}(\Delta^h_{jk} + a^h_{jk}) + g_{hk}a^h_{ij} = 0. \tag{14}$$

Interchanging i and j one has

$$g_{jh}(\Delta^h_{ik} + a^h_{ik}) + g_{hi}(\Delta^h_{kj} + a^h_{kj}) + g_{hk}a^h_{ji} = 0.$$

One-half the sum of these equations is equation (13). One-half their difference is

$$g_{hj}(\Delta^h_{ki} + a^h_{ki}) + g_{hi}(\Delta^h_{jk} + a^h_{jk}) + g_{hk}a^h_{ij} = 0. \tag{15}$$

When to this equation is added the two equations when i, j, k are permuted cyclically twice, the result is

$$g_{hi}(2\Delta^h_{jk} + 3a^h_{jk}) + g_{hj}(2\Delta^h_{ki} + 3a^h_{ki}) + g_{hk}(2\Delta^h_{ij} + 3a^h_{ij}) = 0. \tag{16}$$

This equation is satisfied by equations of the form

$$a^h_{ij} = -\frac{2}{3}\Delta^h_{ij}, \tag{17}$$

as is also equation (15) in consequence of (6). For these values equation (13) is equal to

$$g_{hk}a^{k}_{ij} = \frac{1}{3}(g_{hj}\Delta^{h}_{ik} + g_{hi}\Delta^{h}_{jk}). \tag{18}$$

From (10) and (17) it follows that

$$\Gamma^{h}_{ij} = \Delta^{h}_{ij} + a^{h}_{ij} = \frac{1}{3}\Delta^{h}_{ij}, \tag{19}$$

so that in consequence of (7)

$$\Gamma_i \equiv \Gamma^{j}_{ij} = 0. \tag{20}$$

When equation (11) is multiplied by $g^{iq}g^{pj}$ and i and j are summed, the results after p and q are replaced by i and j respectively is

$$g^{ij}{}_{,k} + g^{hj}\Gamma^{i}_{hk} + g^{ih}\Gamma^{j}_{kh} = 0.$$

When this equation is multiplied by g_{ij} and i and j are summed, the result is

$$(\log w)_{,k} = \Gamma^{h}_{hk}, \tag{21}$$

where w is the square root of minus the determinant of g_{ij}. If we put

$$\mathfrak{g}^{ij} = wg^{ij},$$

we have from the above equation (22)

$$\mathfrak{g}^{ij}{}_{,k} + \mathfrak{g}^{hj}\Gamma^{i}_{hk} + \mathfrak{g}^{ih}\Gamma^{j}_{kh} - \mathfrak{g}^{ij}\Gamma^{h}_{hk} = 0.$$

Denoting by $\mathfrak{g}^{\underset{\sim}{ij}}$ the skew-symmetric part of $\mathfrak{g}^{ij}$, and by $\mathfrak{g}^{\overline{ij}}$ the symmetric part, not to be confused with $\mathfrak{g}^{\underset{\sim}{ij}}$ previously defined, one has

$$\mathfrak{g}^{\underset{\sim}{ij}}{}_{,k} + \mathfrak{g}^{\overline{hj}}\Gamma^{i}_{\underset{\sim}{hk}} - \mathfrak{g}^{\overline{hi}}\Gamma^{j}_{\underset{\sim}{hk}} + \mathfrak{g}^{\underset{\sim}{hj}}\Gamma^{i}_{\underline{hk}} - \mathfrak{g}^{\underset{\sim}{hi}}\Gamma^{j}_{\underline{hk}} - \mathfrak{g}^{\underset{\sim}{ij}}\Gamma^{h}_{\underline{hk}} = 0.$$

In consequence of (20) it follows that

$$\mathfrak{g}^{\underset{\sim}{il}}{}_{,i} = 0. \tag{23}$$

When the condition of integrability is applied to equation (9) it is found[3] that the quantities Γ^{h}_{ijk}, defined by

$$\Gamma^{h}_{ijk} = \Gamma^{h}_{ij,k} - \Gamma^{h}_{ik,j} + \Gamma^{l}_{ij}\Gamma^{h}_{lk} - \Gamma^{l}_{ik}\Gamma^{h}_{lj},$$

are components of a tensor. Contracting for j and k we have in consequence of (20)

$$\Gamma_{ij} = \Gamma^{h}_{ij,h} - \Gamma^{l}_{ih}\Gamma^{h}_{lj} + \Gamma^{l}_{ij}\Gamma^{h}_{\underline{lh}} - \Gamma^{h}_{\underline{ih},j}. \tag{24}$$

The symmetric and skew-symmetric parts of this tensor are

$$\Gamma_{\underline{ij}} = \Gamma^{h}_{\underline{ij},h} - \Gamma^{l}_{\underline{ih}}\Gamma^{h}_{\underline{jl}} + \Gamma^{l}_{\underset{\sim}{ih}}\Gamma^{h}_{\underset{\sim}{jl}} + \Gamma^{l}_{\underline{ij}}\Gamma^{h}_{\underline{lh}} - \frac{1}{2}(\Gamma^{h}_{\underline{ih},j} + \Gamma^{h}_{\underline{jh},i}),$$

$$\Gamma_{ij} = \Gamma^h_{ij,\,h} + \Gamma^l_{ih}\Gamma^h_{jl} - \Gamma^l_{jh}\Gamma^h_{il} + \Gamma^l_{ij}\Gamma^h_{lh} - \frac{1}{2}\left(\Gamma^h_{ih,} - \Gamma^h_{jh,\,i}\right).$$

From (10) and (19) one has

$$\Gamma^h_{ij} = \Delta^h_{ij} + a^h_{ij}, \qquad \Gamma^h_{ij} = \frac{1}{3}\Delta^h_{ij}, \tag{25}$$

which enables one to express Γ_{ij} in terms of g_{ij} and g^{ij}.

The Ricci tensor R_{ij} in terms of g_{ij} is given by[4]

$$R_{ij} = -\Delta^h_{ij,\,h} + \Delta^l_{ih}\Delta^h_{jl} - \Delta^l_{ij}\Delta^h_{lh} + \frac{1}{2}\left(\Delta^h_{ih,} + \Delta^h_{jh,\,i}\right), \tag{26}$$

since

$$\Delta^h_{ih} = (\log u)_{,i}$$

where u is the square-root of minus the determinant of g_{ij}.

By means of (25) and (26) it is found that

$$\Gamma_{ij} = -R_{ij} + a^h_{ij;\,h} - a^l_{ih}a^h_{jl} + a^h_{ij}a^l_{hl} + \frac{1}{9}\Delta^l_{ih}\Delta^h_{jl} - \frac{1}{2}\left(a^h_{ih;\,j} + a^h_{jh;\,i}\right), \tag{27}$$

where a semi-colon followed by index indicates a covariant derivative based upon the g_{ij}.

In like manner

$$\Gamma_{ij} = \frac{1}{3}\left(\Delta^h_{ij;\,h} + \Delta^h_{jl}a^l_{ih} + \Delta^h_{il}a^l_{jh} + \Delta^l_{ij}a^h_{lh}\right) - \frac{1}{2}\left(a^h_{ih;} - a^h_{jh;\,i}\right). \tag{28}$$

In consequence of (21) the tensor Γ_{ij}, defined by (24), is the Einstein tensor R_{ij}.[5] His results are (22), (23), $\Gamma_i = 0$, and

$$R_{ij} = 0, \qquad R_{ij,\,k} + R_{jk,\,i} + R_{ki,\,j} = 0 \tag{29}$$

He interprets (23) as the expression for the vanishing of the magnetic current, and from the tensor Δ_{ijk} he derives the electric current density.

From (27) it follows that, if $\Gamma_{ij} = 0$ the first of equations (29) is

$$R_{ij} = A_{ij}$$

where R_{ij} is the Ricci tensor, not to be confused with the present Einstein tensor R_{ij}, where A_{ij} is given by (27). These quantities must be such that the identities[6]

$$g^{ik}\left(R_{ij;\,k} - \frac{1}{2}R_{ik;\,j}\right) = 0$$

be satisfied.

[1] These PROCEEDINGS, **37,** 311 (1951).

[2] *R. G.*, p. 17. This refers to the author's *Riemannian Geometry*, Princeton University Press.

[3] Eisenhart, L. P., "Non-Riemannian Geometry," *Amer. Math. Soc. Colloquium Publications*, **5,** p. 5.

[4] *R. G.*, p. 21.

[5] Einstein, A., *The Meaning of Relativity*, Fourth Edition, Princeton University Press, pp. 144, 146, 147.

[6] *R. G.*, p. 82.

A THEORY OF FERROMAGNETISM

BY LINUS PAULING

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated April 1, 1953

The properties of ferromagnetic substances are in reasonably good accord with the theory of Weiss.[1] In this theory it is postulated that the atomic magnets tend to be brought into parallel orientation not only by an applied magnetic field but also by an inner field which is proportional to the magnetization of the substance. The inner field is not due to magnetic interaction of the magnetic moments of the molecules, but to electrostatic interactions, which are related to the orientation of the magnetic moments of electrons through the Pauli principle. During the past twenty-five years many efforts have been made to develop a precise theory of the interactions that produce the inner field, and to account in this way for the observed magnetic properties of ferromagnetic substances, but these attempts have not been successful—no one has published a theory of the electronic structure of ferromagnetic substances that permits reasonably good predictions to be made of the values of the saturation magnetic moment and the Curie temperature.

I have now formulated a theory of ferromagnetism that seems to provide a simple explanation of the phenomenon. In the following paragraphs it is shown that the straightforward application of the theory to iron leads, with use only of spectroscopic data for the iron atom, to the predicted values 2.20 Bohr magnetons for the saturation magnetic moment per iron atom and 1350° K for the Curie temperature; these values are in reasonable agreement with the experimental values, 2.22 magnetons and 1043° K, respectively.

The theory can also be applied to cobalt and nickel, and to alloys. In this application, however, a complicating factor must be considered—

the contribution of two different values for the atomic magnetic moment. Satisfactory agreement between theory and experiment is obtained when the ratio of contributions of the two atomic moments has the value assigned to it in a discussion of the electronic structure of these metals that was published fifteen years ago.[2]

The Nature of the Theory.—In discussion of the transition metals the electrons may be divided into two classes, the conduction electrons and the atomic electrons. The conduction electrons are the electrons for which the exchange integrals between adjacent atoms are large, and the atomic electrons are the electrons for which the exchange integrals are small. Some investigators have thought that the electrons occupying $4s$ and $4p$ orbitals comprise the conduction electrons, and those occupying the $3d$ orbitals (plus those in inner shells) comprise the atomic electrons. The first quantum mechanical theory of ferromagnetism is that of Heisenberg.[3] It was shown through the work of Hund, Heitler and London, and others that the exchange integral between electrons with antiparallel spin on the same atom usually has positive sign, whereas that between electrons with antiparallel spin on different atoms has negative sign. In consequence, atomic electrons align their spins parallel, to the extent permitted by Pauli's principle. Moreover, it would be expected that atoms that have a magnetic moment, because of the presence of one or more unpaired electrons on each atom, would tend to align their moments antiparallel. The atomic moments of atoms in a crystal would then tend to divide into two classes, such that the moments of the neighbors of each atom are oriented antiparallel to the moment of that atom. This would give rise to antiferromagnetism (when the moments of the two kinds of atoms are equal) or to ferrimagnetism (when the moments are unequal). Heisenberg assumed, in order to account for the phenomenon of ferromagnetism, that under certain circumstances the exchange integrals of electrons with antiparallel spins on adjacent atoms are positive; the atomic moments would then tend to line up parallel to one another. Heisenberg's theory is unsatisfactory in that there is no independent evidence in support of the assumption that exchange integrals between atoms are ever positive; this remains an *ad hoc* assumption. Moreover, the theory does not provide a reasonable explanation of the observed values of the saturation magnetic moment.

A somewhat different suggestion about the origin of ferromagnetism was made by Bloch.[4] The conduction electrons usually occupy the low levels of a conduction band of orbitals in pairs, with opposed spins. Pauli[5] pointed out that at temperature equilibrium a few of the pairs, occupying the uppermost levels, would be uncoupled by thermal agitation, one of the electrons being raised to a slightly higher level. These uncoupled electrons could then orient themselves independently in a magnetic field; Pauli in

this way explained the temperature-independent paramagnetism observed for the alkali metals and some other metals. Bloch observed that under some circumstances (proper ratio of exchange integrals to the density of states in energy) the exchange energy of the conduction electrons could accomplish some uncoupling, producing a set of conduction electrons occupying separate orbitals, and with parallel spins. The theory has been discussed further by Slater,[6] who has suggested that a rather weak interaction between 3*d* orbitals on adjacent atoms would give rise to a narrow band of levels, with high density of states in energy, and that the conditions for ferromagnetism might be satisfied in this way. In order to account for the observed moment 2.22 magnetons per atom for iron, the assumption is made that 0.22 electron per atom is transferred from the 3*d* orbitals to the 4*s* orbital, constituting the main conduction band; the atomic moment is then ascribed to the hole of 2.22 unoccupied orbitals of the 3*d* subshell, which contains 5.78 electrons.

Recently it was pointed out by Zener[7] that the atomic moments, in parallel orientation, might react with the electrons in the conduction band in such a way as to uncouple some of the pairs, producing a set of conduction electrons occupying individual orbitals, and with spins parallel to the spins of the atomic electrons. Zener assumed that the conduction band for the transition metals is formed by the 4*s* orbitals of the atoms, and that there is somewhat less than one conduction electron per atom in iron, cobalt, and nickel. Like Slater, he attributed the atomic magnetic moments to the partially filled 3*d* subshell.

The new theory is based upon the assumption that there are about six conduction electrons per atom in the transition metals. In the first discussion of the new metallic valences[2] the value 5.78 was assigned as the normal metallic valence of the atoms chromium, manganese, iron, cobalt, and nickel. This value was obtained by subtracting from the total number of outer electrons of iron (eight electrons, outside of the argon shell) the number 2.22 of unpaired electrons, as indicated by the value 2.22 for the saturation magnetic moment of iron. It now seems probable that the valence is equal exactly or nearly exactly to six. In the resonating-valence-bond theory of metals[8] it is assumed that the 3*d*, 4*s*, and 4*p* orbitals of an atom hybridize to form bond orbitals, which are combined with bond orbitals of other atoms in the crystal, with use of Bloch functions, to produce orbitals that are occupied by electron pairs. These orbitals are closely similar to those used for occupancy by single electrons or by two electrons with opposed spins in the usual electron theory of metals. It might be expected that as many atomic orbitals could be combined in this way into orbitals for conduction electrons (bonding electrons) as satisfy the criterion of large overlapping, characteristic of good bond orbitals.[9] The best bond orbital that can be formed from

$3d$, $4s$, and $4p$ orbitals is a $3d^5 4s 4p^3$ orbital, for which the bond strength has value 3.000. Two equivalent orbitals orthogonal to the first one can be constructed. It has been found that a fourth, a fifth, and a sixth of nearly equal strength can also be made, but that a seventh orbital, orthogonal to the other six, is a much poorer bond orbital. Hence it seems not unreasonable to assume that only six bonds will be formed by an atom in a metal, and that the remaining orbitals may be occupied by electrons which interact only weakly with electrons on other atoms.

In iron, cobalt, and nickel there are some unpaired atomic electrons, giving rise to an atomic magnetic moment. There are also six electrons per atom which occupy a conduction band—these are the valence electrons. It is assumed that the direct interaction of the atomic electrons on neighboring atoms is small, and that the exchange integral is, as is customary, negative. If this interaction is larger than the interaction with the conduction electrons, which will be discussed immediately, the moments on adjacent atoms will tend to align themselves in the antiparallel orientation, and the substance will be antiferromagnetic. If the direct interaction energy of the moments is smaller than the possible interaction energy with the conduction electrons, however, the substance will be ferromagnetic.

The conduction electrons occupy metallic orbitals which are formed by linear combination of atomic bond orbitals, which are dsp hybrids. A metallic electron, which may be considered to move from atom to atom through the crystal, interacts with the electrons producing the atomic moment of each atom to the extent that a bond orbital of the atom contributes to the metallic orbital. The exchange integral for the interaction is positive in sign, according to Hund's principle, and the interaction results in stabilization if the spin of the metallic electron is parallel to the atomic moment, and in destabilization if the spin is antiparallel. There is, then, the possibility of stabilization of the metal crystal if the atomic moments are aligned parallel to one another, and a significant number of metallic electrons, occupying separate metallic orbitals, align themselves with spins parallel to the atomic moments. If the energy of interaction of a metallic electron with an atomic moment is $-\epsilon$ for parallel spin and $+\epsilon$ for antiparallel spin, the width of the conduction band, at the top of the doubly occupied conduction levels, that can be occupied by single electrons, with parallel spin, will be equal to 2ϵ. The energy of uncoupling the first pair of electrons (occupying the topmost level) is zero, so that a gain in stabilization through the change in energy -2ϵ will be achieved by uncoupling this pair of electrons. The gain in energy for the last pair is, however, zero, and the average change in energy is $-\epsilon$ per electron pair, or $-\epsilon/2$ per electron. If n conduction electrons are thus uncoupled, the ferromagnetic crystal will be stabilized through this interaction by the

energy change $-n\epsilon/2$. The energy required to reverse the direction of an atomic moment is equal to $2n\epsilon$; accordingly the energy of interaction of the atomic moments and the uncoupled conduction electrons acts in the way corresponding to a Weiss field.

Quantitative calculations can be made on the basis of the assumption that the density of levels in energy for the conduction band is given by the simple expression for the free electron in a box, and the interaction energy ϵ of a *dsp* hybrid conduction electron and the atomic moment can be calculated from the spectroscopic values of the energy of interaction of electrons in the isolated atom. The results of this calculation for iron are discussed in the following section.

Application of the Theory to Iron.—We assume, in essential accordance with earlier conclusions,[2, 8] that six of the eight outer electrons of the iron atom are valence electrons, occupying d^3sp^2 orbitals, and the remaining two are atomic electrons.

The electron configuration of the atom in the metal is accordingly assumed to be $3d^54s4p^2$. This assumption differs considerably from the usual assumption that the electron configuration is close to $3d^8$—for example, Slater's assumption that the configuration is $3d^{7.78}4s^{0.22}$. It is interesting to note, however, that the normal state[10] of the isolated iron atom is not based on the configuration $3d^8$ but on the configuration $3d^64s^2$; its Russell-Saunders symbol is ^5D. The first excited state is $3d^74s\ ^5F$, and the third configuration to appear, with increasing energy, is $3d^64s4p$. There are, in fact, twenty-eight Russell-Saunders states that appear in the FeI spectrum below the first state, $3d^8\ ^3F$, that is assigned to the configuration $3d^8$. The spectroscopic data indicate that the $4s$ atomic orbital is more stable than the $3d$ orbital by about 1.2 ev, and that the energy of a $4p$ electron is only about 1.4 ev greater than that of a $3d$ electron. The configuration $3d^54s4p^2$ accordingly lies about 4 ev above the normal configuration, $3d^64s^2$, and only 1.6 ev above the configuration $3d^8$. The bond energy of six valence electrons can, of course, easily effect the promotion to the assumed configuration $3d^54s4p^2$.

The exchange energy $-\epsilon(4s, 3d)$ of a $4s$ electron and a $3d$ electron with parallel spin can be calculated from the separation of the states $3d^74s\ ^5F$ and $3d^74s\ ^3F$. The separation of these states (averaged over the levels with different values of total moment J) is 4947 cm^{-1}. In the ^5F state the $4s$ electron has spin parallel to the spins of three $3d$ electrons, and in the ^3F state it has spin antiparallel to the spins of these three $3d$ electrons. The value of the exchange energy per electron is accordingly $-4947/6 = -825$ cm^{-1}. For $-\epsilon(3d, 3d)$ one sixth of the difference in energy of $3d^64s^2\ ^5D$ and $3d^64s^2\ ^3D$ was used; its value is -4826 cm^{-1}. The evaluation of $-\epsilon(4p, 3d)$ as -470 cm^{-1} is similar, although somewhat less straightforward, because it involves averaging over a number of Russell-

Saunders states. Assuming that the valence electrons at the top of the band have the average hybrid character $3d^34s4p^2$, the interaction energy of one of these valence electrons and an atomic electron, assumed to be approximately a $3d$ electron, is found to be -2707 cm^{-1}, or -0.334 ev, with probable error about 10%.

The interaction energy of the valence electron with the two atomic $3d$ electrons, with parallel spins, is accordingly -0.67 ev, and the width of the energy band that would be occupied by uncoupled valence electrons is 1.34 ev. The number of orbitals in this band can be calculated from the equation for the distribution of energy levels for an electron in a box. The number of levels per atom is

$$n_0 = 4\pi V (2m)^{3/2} E^{1/2}/3h^3,$$

where V is the atomic volume, m is the mass of the electron, E is the kinetic energy of the electron, and h is Planck's constant. For iron, with $V = 11.72 \times 10^{-24}$ cm^3, and six electrons per atom, which can occupy three orbitals, in pairs, the value of E, the energy of the top level occupied by conduction electrons if they are all paired, is calculated to be 23.3 ev. Differentiation of this equation leads to the equation

$$dn = 2\pi V (2n)^{3/2} E^{1/2} dE/h^3$$

as the number of levels dn with the width in energy dE. Substitution of the value 1.34 ev for dE leads to the value 0.26 for dn. Accordingly we predict that 0.26 of the conduction electrons will be uncoupled by interaction with the atomic moments, and will align themselves parallel to these moments.

The experimental values of the gyromagnetic ratio and the observed magnetic moments for paramagnetic compounds of transition metals show that the orbital moment is nearly completely quenched. On multiplying 2.26 by the ratio 1.946/2.000 of the gyromagnetic ratio to the spin value, the predicted value of the saturation magnetic moment per atom for iron is obtained as 2.20 magnetons, in excellent agreement with the experimental value, 2.22 magnetons.

The calculated energy of interaction of an atomic moment and the Weiss field (0.26 uncoupled conduction electrons per atom) for magnetic saturation is 0.135 ev, or 3070 cal. mole^{-1}. According to the Weiss theory the Curie temperature is equal to this energy of interaction divided by $3k$, where k is Boltzmann's constant. The effect of spatial quantization of the atomic moment, with spin quantum number S, is to introduce the factor $(S + 1)/S$; that is, the Curie temperature is equal to $n\epsilon(S + 1)/3Sk$. For iron, with $S = 1$, the predicted value for the Curie constant is 1350°K, in rough agreement with the experimental value, 1043°K.

Cobalt and Nickel.—The discussion of the magnetic properties of cobalt

and nickel is somewhat more complicated than that for iron, because of the necessity of giving consideration to the metallic orbital. It has been pointed out[8] that on the basis of the resonating-valence-bond theory of metals a metallic structure, showing large electronic conductivity, can occur only if many of the atoms in the metal have available one orbital not ordinarily occupied by an electron; this orbital can act as receptor for an additional electron, permitting the unsynchronized resonance of the shared electron pairs in the metal. When the metallic orbital is occupied by an electron the atom has a negative charge, and another atom in the metal has a positive charge; it is the motion of these negative charges and positive charges through the metal that gives rise to electronic conduction. (A discussion of the relation between metallic conductivity and negatively charged and positively charged atoms in the metal on the basis of the conventional electronic theory of metals has been given by Slater.)[11] An atom of cobalt, with six valence electrons and three atomic electrons, essentially $3d$ electrons, can have the spins of the atomic electrons parallel by utilizing all of its orbitals; a metallic orbital is made available only if two of the atomic electrons have opposed spins. The metal can be described as having some atoms with resultant spin quantum number $3/2$ and some atoms with spin $1/2$. In the same way the atom of nickel can be described as having some atoms, without a metallic orbital, with spin 1 and some atoms, with a metallic orbital, with spin 0. The number of atoms without a metallic orbital has been estimated[8] to be 28%. In the following calculations we use the value 28.5%, which is obtained from the value 28.57 for the electron number that marks the upper boundary of the ferromagnetic region, as given by the intercept of the curve of values of saturation magnetic moment corrected by extrapolation for isoelectronic sequences to the value 0°K for the Curie temperature (Taglang).[12]

With this assumption and use of the value 1.94 for the g-factor the average magnetic moment per atom of cobalt is calculated to be 1.52 magnetons. The spectroscopic data for cobalt do not provide very reliable values for the interaction energies of $3d$, $4s$, and $4p$ electrons. The data for chromium and manganese, when compared with those for iron, indicate that the interaction energies increase by 5% with unit increase in the atomic number. We accordingly use the value 0.70 ev for ϵ, which leads to 0.20 uncoupled valence electrons per atom. The predicted saturation magnetic moment for cobalt is thus 1.72 magnetons, in good agreement with the experimental value 1.71 magnetons. The value of the Curie temperature is calculated with use of the average value of the magnetic moment per atom and the average value of $(J + 1)/J$ to be 1190°K, which is somewhat smaller than the experimental value 1393° K.

In nickel 28.5% of the atoms, without a metallic orbital, have two atomic electrons with parallel spins, and magnetic moment 1.94 (corre-

sponding to $J = 1$, with $g = 1.94$). The value of the interaction energy ϵ is 0.74 ev, and the calculated number of uncoupled valence electrons per atom is 0.064. The saturation magnetic moment per atom is thus calculated to be 0.617, in good agreement with the observed value 0.606 magnetons. The calculated value of the Curie temperature, 367° K, is considerably smaller than the observed value, 631° K.

For alloys of iron, cobalt, nickel, and copper the calculated values of saturation magnetic moments agree closely with the observed values; in particular, the maximum value of about 2.48 magnetons at electron number about 26.3 is reproduced by the theory. There is, however, only rough agreement between the observed and calculated values of the Curie temperature.

An obvious refinement of the simple theory for cobalt and nickel and their alloys can be made which leads to a significant increase in the calculated value of the Curie temperature. The foregoing calculation for nickel, for example, is based upon the assumption that the uncoupled valence electrons spend equal amounts of time on the nickel atoms with $J = 1$ and the nickel atoms with $J = 0$. However, the stabilizing interaction of the spins of the valence electrons and the parallel atomic moments would cause an increase in the wave function for the valence electrons in the neighborhood of the atoms with $J = 1$ and the parallel orientation. This effect also produces a change in the shape of the curve of saturation magnetization as a function of temperature. The details of this refined theory will be published later.

The simple theory accounts for the deviation of the slope of the curve of saturation magnetic moment against average electron number from the values $+1$ and -1 by about 10% as resulting from an induced moment contributed by the conduction electrons. This explanation was suggested and given empirical support by Forrer[13] and Taglang.[12]

Gadolinium.—For gadolinium the saturation magnetic moment is 7.12 magnetons and the Curie temperature is 290° K.[14] There is little doubt that three of the ten outer electrons of gadolinium are valence electrons and the other seven are atomic electrons occupying $4f$ orbitals, with spectroscopic state ^8S. These seven atomic electrons would give rise to the magnetic moment 7.00 magnetons, and the small increment may be attributed to uncoupled valence electrons.

Spectroscopic data are not available to permit an independent estimate of the magnitude of the interaction energy of a valence electron and the seven atomic electrons. If we assume the value 0.31 ev for this quantity, ϵ, the calculated value of the Curie temperature, 290° K, is equal to the observed value, and the calculated value of the saturation magnetic moment, 7.19 magnetons, is slightly greater than the observed value, 7.12. The small disagreement might be explained as resulting from a small

contribution of orbital moment, reducing the g-factor from 2 to 1.98; the value 1.95 ± 0.03 for the g-factor has been reported from microwave resonance absorption measurements of gadolinium metal in the paramagnetic region.[15]

Other Interactions.—In the foregoing calculations the direct exchange interactions of the atomic electrons on neighboring atoms have been neglected. These exchange interactions are expected in general to tend to cause the atomic moments on adjacent atoms to align themselves in the antiparallel orientation. If these interactions and other interactions (such as those involving an unsymmetrical distribution of the electrons in an electron-pair bond between adjacent atoms) are less important than the interactions that stabilize the parallel orientation of the atomic moments the substance will be ferromagnetic at low temperatures, and the value of the saturation magnetic moment will be that calculated by the above methods. However, the neglected interactions, although they are not expected to change significantly the saturation magnetic moment, could have a large effect on the calculation of the Curie temperature.

It has been pointed out by Bozorth[16] that not all of the alloys of the iron-group transition metals have ferromagnetic saturation moments that correspond to the simple theory of six valence electrons and 0.715 metallic orbital per atom. It is likely that the exceptional alloys show ferrimagnetism or antiferromagnetism—that in these alloys the interactions that stabilize antiparallel orientation of the moments of adjacent atoms are stronger than those favoring parallel orientation. Until these interactions have been subjected to quantitative treatment the theory of ferromagnetism, ferrimagnetism, and antiferromagnetism will remain incomplete.

* Contribution No. 1791. This investigation was carried out in connection with a program of research on metals supported by contract N6onr-24432 between the California Institute of Technology and the Office of Naval Research.

[1] Weiss, P., *J. de Physique*, **6**, 667 (1907); **1**, 166 (1930).

[2] Pauling, L., *Phys. Rev.*, **54**, 899 (1938).

[3] Heisenberg, W., *Z. f. Physik*, **49**, 619 (1928).

[4] Bloch, F., *Z. f. Physik*, **57**, 545 (1929).

[5] Pauli, W., *Z. f. Physik*, **41**, 81 (1927).

[6] Slater, J. C., *Phys. Rev.*, **36**, 57 (1930); **49**, 537, 931 (1936); **52**, 198 (1937).

[7] Zener, C., *Phys. Rev.*, **81**, 440 (1951).

[8] Pauling, L., *Proc. Roy. Soc. London*, **A 196**, 343 (1949).

[9] Pauling, L., *J. Am. Chem. Soc.*, **53**, 1367 (1931).

[10] Moore, C. E., *Atomic Energy Levels*, Vol. I, June 15, 1949, and Vol. II, August 15, 1952, U. S. Government Printing Office, Washington 25, D. C.

[11] Slater, J. C., *Phys. Rev.*, **35**, 509 (1930).

[12] Taglang, P., *Compt. rend.*, **229**, 704, 994 (1949).

[13] Forrer, R., *J. de Physique*, **10**, 181 (1939); **1**, 74 (1940).

[14] Urbain, G., Weiss, P., and Trombe, F., *Compt. rend.*, **200**, 2132 (1935); Trombe, F., *Ann. de Physique*, **7**, 385 (1937).

[15] Kip, A. F., Kittel, C., Portis, A. M., Barton, R., and Spedding, F. H., *Phys. Rev.*, **89**, 518 (1953).
[16] Bozorth, R. M., *Phys. Rev.*, **79**, 887 (1950).

EQUAL RATIOS OF ENERGIES OF NUCLEAR EXCITED STATES OF EVEN-EVEN ISOTOPES

By Martin G. Redlich*

Palmer Physical Laboratory, Princeton University

Communicated by E. P. Wigner, April 15, 1953

The purpose of this note is to point out a regularity in the measured energies of some excited states of isotopic nuclei with even numbers both of neutrons and protons (N and Z). The experimental data indicate that although the addition of two or four neutrons to an even-even nucleus changes the energies of its excited states, in most instances it does not alter the *ratio* between the energies of its first and second excited states.

In order to describe this relation, as well as some similar ones, let $E_m(A)$ and $E_n(A')$ denote the energies of the mth and nth excited states of two even-even isotopes with mass numbers A and $A' = A + 2$ or $A + 4$, and further, let

$$r_m(A) = \frac{E_m(A)}{E_1(A)}.$$

In this note, if $r_n(A')$ is equal within 3% to $r_m(A)$, the nth level of isotope A' will be said to *correspond* to the mth level of isotope A. This correspondence will be expressed symbolically as $n \sim m$. (The correspondence is not symmetric.) For six of the nine pairs of isotopes on which data are available, $2 \sim 2$. The distribution of correspondences among the remaining pairs is: $2 \sim 3$ and $3 \sim 4$ for one pair; $3 \sim 2$ for another. There is only one pair, Pb^{204} and Pb^{208}, for which no correspondence is observed. It is possible to conclude from experimental evidence which will be discussed in a subsequent paragraph that the spins of the second excited states of these Pb isotopes are not equal. In view of this fact, the absence of a correspondence between these states might be expected. On a graph of $E(A)$ vs. $E(A')$ the correspondence $2 \sim 2$ would entail that the points $(E_1(A), E_1(A'))$ and $(E_2(A), E_2(A'))$ lie on a straight line through the origin.

The experimental data are summarized in table 1. The energies of at least two excited states of many even-even isotopes can be obtained from decay schemes. All elements with two or more such isotopes whose levels were found in a recent paper[1] or in the new Table of Isotopes[2] are listed in

TABLE 1

ENERGIES OF EXCITED STATES OF EVEN-EVEN ISOTOPES

The number below each datum is its estimated uncertainty times 10^3. It is enclosed in parentheses unless based upon estimates made in the original papers.

ELEMENT AND Z	A	N	$E_1(A)$ IN MEV	$r_2(A) = \frac{E_2(A)}{E_1(A)}$	$r_3(A) = \frac{E_3(A)}{E_1(A)}$	$r_4(A) = \frac{E_4(A)}{E_1(A)}$	$\frac{r_2(A)}{r_2(A_{\min.})}$	OTHER CORRESPONDENCES
Ti	46	24	0.890[a]	2.258				
			5	13				
22	48	26	0.990[b]	2.242	2.333		0.993	
			(2)	(25)	(6)		(17)	
Mo	94	52	0.874	3.124	3.741			$r_3(96)/r_2(94) = 1.005$
			(9)	(43)	(66)			(28)
42	96	54	0.771	2.088	3.139			
			(7)	(28)	(45)			
								$r_2(114)/r_3(110) = 0.979$
Cd	110	62	0.655	2.159	2.348	3.394		4
			1	5	5	(20)		
48	114	66	0.556	2.299	3.338			$r_3(114)/r_4(110) = 0.984$
			1	4	10			9
Sn	116	66	1.274	1.641	1.852			
			6	16	8			
50	120	70	1.30[a]	1.692			1.031	
			30	39			33	
Te	122	70	0.568	2.183				
			(10)	(38)				
52	124	72	0.607	2.203	3.278		1.009	
			(6)	(24)	(46)		(31)	
Ba	134	78	0.793	1.723	1.759	2.467[c]		
			2	4	5	7		
56	138	82	1.44	1.681[d]	2.000		0.976	
			(30)	(28)	(42)		(19)	
Pt	192	114	0.316[e]	1.937[e]	2.479[e]	2.910[e]		
			<1	<1	<1	1		
78	194	116	0.328[b]	1.887[d]	6.402		0.974	
			2	12	68		6	
	196	118	0.358[a]	1.922			0.992	
			20	136			71	
Pb	204	122	0.374	3.420				
			(2)	(18)				
82	208	126	2.620	1.222	1.328	1.523[d]		
			(8)	(1)	(2)	(3)		

[a] The energy of this state is either E_1 or $E_2 - E_1$. The nucleus in the second excited state decays by emission of two γ-rays in cascade; their sequence is not definitely known.

[b] The position of this level is not definitely established.

[c] There may exist another level between the ones denoted 3rd and 4th levels here.

[d] The position of the higher level is not definitely established.

[e] The decay scheme is not certain.

column 1 of table 1. States observed only in heavy particle reactions were not considered because in the many different possible reactions various sets of states are excited; in addition, the data are generally less accurate. The energy $E_1(A)$ and the ratios $r_n(A)$ for $n = 2$, 3, and 4, together with their estimated uncertainties, are listed in columns 4 to 7. A blank space in the columns for $r_3(A)$ and $r_4(A)$ or for $r_4(A)$ alone indicates that only two or three excited states are known from decay schemes. Footnotes in the table refer to all those levels about whose position in the decay schemes there remains some doubt. It must be emphasized, however, that the decay scheme of no nucleus in table 1 is entirely certain. For example, new levels with energies below that of the highest level listed may be discovered. This seems especially likely for Pb^{204}.

It is plain from columns 5 to 7 that correspondences exist. Quotients of the ratios for two isotopes, $r_n(A)/r_m(A_{\min.})$, are listed in columns 8 and 9 for the four types of correspondence, $n \sim m$, which appear. $A_{\min.}$ is the mass number of the lightest isotope.

The data available (reference 1) are consistent with an assignment of spin and parity $2+$ to the first excited states of about half of the nuclei in the table; there is no information about the remaining ones. There are data bearing on these quantities for higher excited states of both isotopes of two of the elements in table 1:

(1) Ti. γ–γ angular correlation measurements are consistent with an assumption of spins 2 and 4 for the first and second excited states of Ti^{46}. Similarly, such measurements are consistent with spins 2 and 4 for the first and third (until recently believed to be the second) excited states of Ti^{48}. One might expect, therefore, that $3 \sim 2$ instead of $2 \sim 2$, but this does not prove to be the case. There is no conclusive evidence, however, about the spins and parities of any of the states of Ti^{48}.

(2) Pb. The second excited state of Pb^{204} has been assigned spin and parity $7-$. It is an isomeric state with a half-life of about one hour. The mass number of this isotope is classified as "probable" in reference 2. The second excited state of Pb^{208} was assigned spin 4 from γ–γ angular correlation measurements. It is not surprising, therefore, that these two states do not correspond.

It is readily seen that the probability for chance occurrence of the $2 \sim 2$ correspondences is small. Let Q equal the smaller divided by the larger of the two numbers $r_2(A)$ and $r_2(B)$ for two even-even nuclei, not necessarily isotopes, with mass numbers A and B. By definition, $0 < Q \leq 1$. Q is actually restricted to a smaller interval, however, since the spacing of the first two levels is qualitatively similar for different even-even nuclei. The ratios $r_2(A)$ in column 5 of the table, for example, vary only from 1.222 to 3.420, and values of Q for all pairs lie between 0.36 and 1. In the probability calculation an even smaller interval, $0.5 \leq Q \leq 1$, will be taken as the statistical sample space.

Let us select now the quotients Q for the nine pairs of isotopes, with mass numbers A and $A_{\min.}$, on which data are available. Six of these quotients are in the interval $0.97 \leq Q \leq 1$. The probability that six or more among nine numbers chosen at random from a sample space of length 0.50 lie in an interval of length 0.03, i.e., in 6% of the total interval, is

$$\sum_{k=6}^{9} \binom{9}{k} (0.06)^k (0.94)^{9-k} = 3.3 \cdot 10^{-6}.$$

Acknowledgment.—It is with gratitude that I acknowledge Professor E. P. Wigner's advice and interest.

* Hercules Fellow.

[1] Scharff-Goldhaber, Gertrude, *Phys. Rev.*, 1953 (in press).

[2] Hollander, Perlman, and Seaborg, *Revs. Mod. Phys.*, 1953 (in press).

---

*THE EFFECT UPON HEART RATE CONDITIONING OF RANDOMLY VARYING THE INTERVAL BETWEEN CONDITIONED AND UNCONDITIONED STIMULI**

BY P. J. BERSH, W. N. SCHOENFELD, AND J. M. NOTTERMAN

COLUMBIA UNIVERSITY

Communicated by C. H. Graham, April 15, 1953

The experiment reported here continues the investigation of variables which affect the conditioning and extinction of a heart rate response in human beings.[1-3] No record has been found in the literature of any attempt to carry out Pavlovian conditioning by randomly varying the interval between conditioned stimulus (CS) and unconditioned stimulus (US), while providing reinforcement on a regular schedule. The present experiment compares the response strength generated by such a procedure, with that produced by regular and irregular reinforcement schedules involving a constant CS-US interval.

Subject and Apparatus.—Thirty male university students, ranging in age from 20 to 35 years, were used as subjects. After screening for cardiac and psychiatric history,[1] they were assigned at random to the following three groups consisting of ten subjects each: (*a*) Group C_r: constant CS-US interval, regular reinforcement; (*b*) Group C_i: constant CS-US interval, irregular reinforcement; and (*c*) Group V_r: randomly varying CS-US intervals, regular reinforcement.

The apparatus was the same as that employed in the earlier studies and is described in detail elsewhere.[1] The important features were an electro-

cardiograph; a 750-cps tone, at 20 db, transmitted to the subject through a headset; and a 30-v. a. c. shock administered through two electrodes strapped to the palmar and dorsal sides of the subject's left hand.

Procedure.—Each subject was carried through three experimental phases in a single session: basal (preconditioning); conditioning; and extinction. Before the start of the basal phase, all subjects were given the following instructions:

"This is an experiment to determine your heart rate reaction to tone, and your heart rate reaction to electric shock. You will hear a series of tones through this headset which will soon be placed over your ears. A little later in the experiment, you will receive a series of electric shocks administered to your left hand; you will not be shocked through the electrodes I am now placing on your chest and left leg. These electrodes enable us to record your heart action. The shocks you will receive will be strong enough for you to feel them, but will not be strong enough to do you any harm. Remember that they will just pass through your hand, and not your body. Please try not to make any unnecessary movements. I will be keeping a continuous record of your heart action while you are in here, and since the apparatus is necessarily very sensitive, your moving around too much would interfere with the recording."

During the basal and extinction phases, all subjects were exposed to uniform training. The basal phase consisted of 20 tone-alone, and the extinction phase of 10 tone-alone, trials. On each trial, the subject's heart rate was recorded continuously from about 30 seconds before tone presentation to about 45 seconds after termination of the tone. On reinforced conditioning trials, recording lasted from 30 seconds before tone presentation to 30 seconds after shock termination. The durations of tone and shock were 1 second and 6 seconds, respectively, with the interval between tone and shock as indicated below. Throughout all phases of the experiment, spacing of trials was randomly varied over a range of from 1–2 minutes.

The procedure during the conditioning phase differed for the three groups as follows:

(*a*) Group C_r. This group received 15 tone-shock pairings. On each trial, shock followed termination of the tone by 9 seconds.

(*b*) Group C_t. This group also was given 15 reinforced (tone-shock) trials, with 10 unreinforced (tone-alone) trials interspersed among them. The first five trials were reinforced.[4] During the subsequent 20 trials, 10 unreinforced presentations of the tone were randomly interspersed with 10 reinforced presentations. On each reinforced trial, the tone-shock interval was 9 seconds, as in the case of Group C_r.

(*c*) Group V_r. This group received 15 consecutive tone-shock trials, with the tone-shock intervals randomized by blocks of five trials each. Within each block, five values of tone-shock interval, namely 4, 6, 9, 12 and

15 seconds, were presented in random sequence. The mean of the randomized intervals was 9.2 seconds, a fact which led to the use of the 9-sec. interval, rather than the 6-sec. interval previously reported,[1,2] on all reinforced trials for the other two groups. While insertion of a 3-sec. instead of a 4-sec. interval would have resulted in a mean interval of exactly 9 seconds for Group V_r, 3 seconds are insufficient to permit at least two heart cycles between tone and shock in the case of all subjects, while a single cycle does not provide a reliable measure of heart rate.

Results.—In the results reported below, the basic data consist of pretone and posttone rates in beats per minute (b/m). The pretone rate was in all cases measured from the two beats immediately preceding tone presentation. For Groups C_r and C_t on all trials, and for Group V_r during the basal

TABLE 1

GROUP MEAN PRETONE AND GROUP MEAN POSTTONE HEART RATE (B/M) DURING EACH PHASE OF THE EXPERIMENT

	PHASE											
	BASAL		CONDITIONING				EXTINCTION					
			ALL TRIALS		LAST FIVE		ALL TRIALS		FIRST FIVE		LAST FIVE	
GROUP	PRE	POST	PRE	POST	PRE	POST	PRE	POST	PRE	POST	PRE	POST
C_r	75.0	75.1	75.5	71.5	76.4	71.1	75.4	73.8	75.3	74.2	75.5	73.4
	$P > 0.05$		$P < 0.001$		$P < 0.01$		$P > 0.05$		$P > 0.05$		$P > 0.05$	
C	79.4	78.5	80.0	73.6	79.3	70.6	79.3	73.4	78.7	72.9	79.8	73.9
	$P > 0.05$		$P < 0.001$		$P < 0.001$		$P < 0.001$		$P < 0.01$		$P < 0.001$	
V_r	80.7	81.7	80.7	77.8	81.7	77.0	80.7	76.7	82.0	76.3	79.3	76.9
	$P > 0.05$		$P < 0.01$		$P < 0.01$		$P < 0.01$		$P < 0.01$		$0.03 > P > 0.02$	

TABLE 2

TESTS OF SIGNIFICANCE BASED UPON PHASE-BY-PHASE INTERGROUP COMPARISONS

	PHASE					
	BASAL	CONDITIONING		EXTINCTION		
COMPARISON		ALL	LAST 5	ALL	FIRST 5	LAST 5
C_t vs. C_r	$P > 0.05$	$P < 0.05$	$P < 0.05$	$P < 0.05$	$P < 0.05$	$P < 0.02$
C_t vs. V_r	$P > 0.05$	$P < 0.01$	$P < 0.05$	$P > 0.05$	$P > 0.05$	$P < 0.02$
V_r vs. C_r	$P > 0.05$	$P > 0.05$	$P > 0.05$	$P > 0.05$	$P < 0.05$	$P > 0.05$

The underlined values are based upon a one-tail t-test applied without the usual parametric assumptions.

and extinction phases, the posttone rate was determined from the final two beats in the 9-sec. interval following termination of the tone. However, on conditioning trials for Group V_r, the posttone rate was based upon the last two beats before onset of shock. Since shock occurred at intervals ranging from 4 to 15 seconds after termination of the tone, the two beats whose measurement was recorded as the posttone value varied in temporal position accordingly. The consequences of this variation in the time of measurement of posttone rate for Group V_r are discussed later.

The phase-by-phase results for the three groups are summarized in table 1. The two values for each phase are the group mean pretone and the group mean posttone heart rates, respectively, in beats per minute. Each

value was obtained by averaging the pretone and posttone means for individual subjects in each group on the indicated trials of the given phase; the basal values represent trials 11–20 of that phase.[1] All intragroup comparisons reported in table 1, or elsewhere in the text, are based upon Pitman's variation of a non-parametric test for paired items developed by Fisher.[5] The P values in table 1 are for the obtained differences between mean pretone and mean posttone rates during each of the phases reported.

For intergroup comparisons, a summary of P values based largely upon Wilcoxon's[6] non-parametric test for unpaired items, is given in table 2.

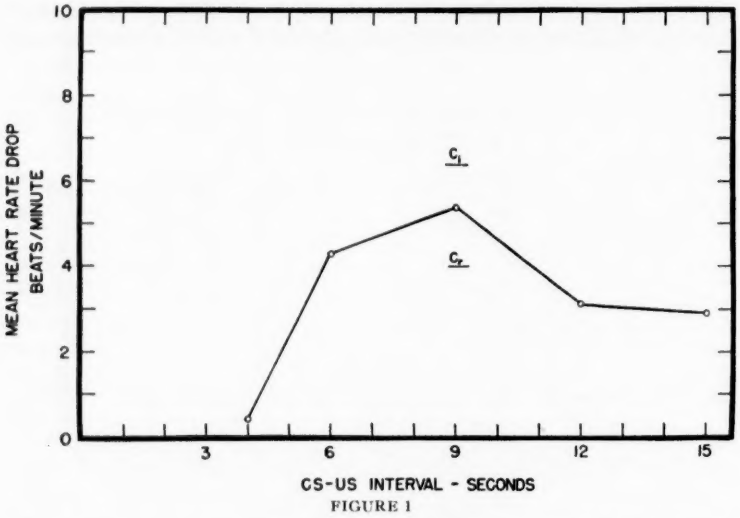


FIGURE 1

Acquisition. Mean heart rate drop in Group V_r as a function of CS-US interval. Means are based on all trials of a given interval occurring during conditioning. The short horizontal lines are the mean heart rate drops (CS-US interval of nine seconds) for Groups C_t and C_r, respectively.

In each case where a significant result is reported, the group mentioned first in the comparison showed the larger heart rate change. Where Wilcoxon's test did not yield significance, the more powerful but less convenient t-test was resorted to without the usual parametric assumptions,[7] the underlined figures being one-tail values; where the value reads simply "$P < 0.05$", the given comparison would not be significant on the basis of a two-tail test. A table entry of "$P > 0.05$" indicates that neither test led to a significant result.

Since the members of Group V_r were exposed to five different CS-US intervals during conditioning, an analysis of heart rate changes was made

separately for each of the intervals (the posttone measure in each case representing the last two beats before shock), and the mean results by intervals for the conditioning phase taken as a whole are graphed in figure 1. Each point is the mean of all determinations for the given interval. In view of the conditioning history of the members of Group V_r, an analysis by intervals also seemed desirable for extinction, and was actually feasible, since the absence of shock during extinction made it possible to obtain a measurement, on every trial, for each of the five intervals previously used dur-

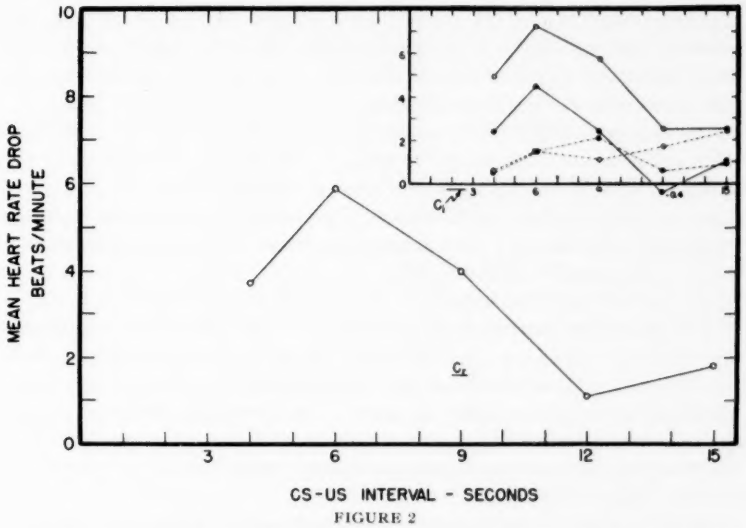


FIGURE 2

Extinction. Mean heart rate drop for Group V_r as a function of CS-US interval. Means are based on all trials of a given interval occurring during extinction. The short horizontal lines are the mean heart rate drops (CS-US interval of nine seconds) for Groups C_r and C_l, respectively. The inset provides analogous data for Groups V_r and C_r for the two halves of extinction. The solid curves are data for Group V_r; the broken curves are data for Group C_r. The open circles represent means based on trials 2–6; the solid circles are means based on trials 7–11.

ing conditioning. In figure 2 the results for this analysis are plotted for the entire extinction phase. The inset represents a similar analysis for the extinction phase divided into halves to correspond with the breakdown for this phase in table 1. Also included in the figure 2 inset are plots of analogous data for Group C_r. These provide an interval-by-interval comparison of the amplitude of the acquired heart rate response during each half of extinction for the two groups previously given shock on every conditioning trial.

Discussion. (*1*) *Acquisition.*—The data in table 1 show that conditioning was successful for Group V_r, as well as for the other two groups. The overall mean drops were 6.4, 4.0 and 2.9 b/m for Groups C_t, C_r and V_r, respectively. While these values point to some superiority of a regular reinforcement schedule for a constant interval (Group C_r) over a varying interval (Group V_r), this is largely due to the procedure for computing the posttone rate for Group V_r. As noted earlier, this rate was measured from the two beats before shock onset, which, from trial to trial, varied from 4 to 15 seconds following the end of the tone. For conditioning taken as a whole, the maximum amplitude of the conditioned response (CR) occurs at the 9-sec. posttone interval (Fig. 1). When the mean change in heart rate is computed for Group V_r solely on the basis of 9-sec. trials, the value becomes 5.4 b/m, indicating a conditioned response at least equal to that for Group C_r.

The strongest conditioning was manifested by Group C_t, as evidenced by the significance tests reported in table 2. The table 2 comparisons between Groups C_t and V_r are biased toward significance by the previously noted method of measuring the Group V_r posttone rate; but even when the mean values for Group V_r are obtained for 9-sec. trials alone, they remain smaller than those for Group C_t ($P \cong 0.05$).

Of some interest is the influence of the varying interval schedule upon the way in which the heart rate varies within each trial as a function of time since the end of the tone. As mentioned above, for conditioning as a whole the CR tends to reach its maximum amplitude at about 9 seconds, the approximate mean of the intervals used. The data also contain some statistically non-significant indications that, as conditioning progresses, there is a temporal shift in the maximum amplitude of the CR so that it tends to occur later and later in the posttone period.

(*2*) *Extinction.*—Comparisons of the relative response strength generated by the three conditioning procedures are also afforded by the extinction data. Table 1 shows that the varying interval schedule generates a significant resistance to extinction. Over the first half of extinction, the acquired response holds up about as well for Group V_r as for Group C_t, and significantly better (*vide* Table 2) for Group V_r than for Group C_r. Furthermore, the mean conditioned response for Group V_r remains significant for the last half of extinction, though, as seen in table 2, there is a significant difference in favor of Group C_t for this part of the extinction phase. On the other hand, none of the heart rate changes manifested by Group C_r during extinction are significant. Presumably extinction occurs very rapidly for Group C_r, since at the end of conditioning (extinction trial 1) the mean heart rate change (a drop of 7.5 b/m) is significant ($P < 0.01$), while on the subsequent extinction trial, the mean drop in rate (2.9 b/m) already is not significant. It should be remembered that the heart rate change on trial 1 of extinction is recorded before the first omission of shock, and that

the effect of such omission is felt only during the second trial. By comparison with a previous study,[1] it appears that a constant 9-sec. interval is less effective than a constant 6-sec. interval where the schedule of reinforcement is regular. Apparently no analogous decrease in effectiveness accompanies the use of a constant 9-sec. instead of a constant 6-sec. interval where the schedule of reinforcement is irregular. This suggests the possibility of a relationship between schedule of reinforcement and the duration of a fixed CS-US interval.

For all groups, the extinction values presented in table 2 are based upon posttone rates taken from the last two beats in the 9-second posttone interval. When heart rate changes during extinction are plotted as a function of the intervals employed in the conditioning of Group V_r, it becomes clear (*vide* Fig. 2 and inset) that the CR is greater for Group V_r than for Group C_r through most of the range of intervals represented. This is especially true for the first half of extinction. During the last half, differences in amplitude still exist at 4 and 6 seconds, though the progress of extinction has about equalized the CR for the 9–15-sec. intervals.

Cross-comparisons between Groups C_r and V_r for intervals other than the 9-sec. interval already reported in table 2, reveal (Wilcoxon test) that for the first half of extinction, the 4-sec. values are also significantly different ($0.03 > P > 0.02$). None of the cross-comparisons for the last half of extinction is significant.

Further support for the finding that the conditioned response shows more resistance to extinction for Group V_r than for Group C_r is afforded by intragroup comparisons involving the posttone heart rates measured at the 4-, 6-, 12- and 15-sec. intervals. (Similar comparisons for the 9-sec. interval are found in table 1.) Thus, for the first half of extinction, Group V_r shows a significant heart rate depression at the 4- ($P < 0.001$), 6- ($P < 0.001$) and 15- ($0.03 > P > 0.02$) sec. intervals on the basis of comparisons of subject mean pretone rates with subject mean posttone rates; for the second half of extinction, the depression is significant at the 6-sec. interval ($0.02 > P > 0.01$). If the heart rate changes manifested by Group V_r during the first half of the extinction are compared with the changes during trials 11–20 of the basal phase, the depression becomes significant for the 12-sec. interval also ($0.02 > P > 0.01$). None of these comparisons reveals any significant heart rate depressions for Group C_r.

Summary.—A procedure of randomly varying a tone-shock interval was successful in conditioning a depression in heart rate in human subjects. The magnitude of the CR obtained with the varying interval group was at least equal to that for a constant interval group, where both groups were regularly reinforced. A constant interval-irregularly reinforced group reached a stronger CR, however, than did the former two groups. The resistance to extinction of the CR for the varying interval group was greater

than that for the constant interval-regular reinforcement group, but less than that for the constant interval-irregular reinforcement group. During the first half of extinction, the CR held up about equally well for the varying interval and the constant interval-irregular reinforcement groups.

* This study was supported by funds provided under contract AF 18(600)-69 with the USAF School of Aviation Medicine, Randolph Field, Texas.

[1] Notterman, J. M., Schoenfeld, W. N., and Bersh, P. J., "Conditioned Heart Rate Response in Human Beings During Experimental Anxiety," *J. Comp. Physiol. Psychol.*, **45**, 1–8 (1952).

[2] Notterman, J. M., Schoenfeld, W. N., and Bersh, P. J., "Partial Reinforcement and Conditioned Heart Rate Response in Human Subjects," *Science*, **115**, 77–79 (1952).

[3] Notterman, J. M., Schoenfeld, W. N., and Bersh, P. J., "A Comparison of Three Extinction Procedures Following Heart Rate Conditioning," *J. Abnorm. Soc. Psychol.*, **47**, 674–677 (1952).

[4] Jenkins, W. O., and Stanley, J. C., Jr., "Partial Reinforcement: A Review and Critique," *Psychol. Bull.*, **47**, 193–234 (1950).

[5] Pitman, E. J. G., *Notes on Non-Parametric Statistical Inference*, Dept. Mathematical Statistics, Columbia Univ., New York, 1948.

[6] Wilcoxon, F., "Individual Comparisons by Ranking Methods," *Biometric Bull.*, **1**, 80–83 (1945).

[7] Moses, Lincoln E., "Non-Parametric Statistics for Psychological Research," *Psychol. Bull.*, **49**, 122–143 (1952).

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

REFERENCES to literature, numbered consecutively, will be placed at the end of the article and short footnotes should be avoided. It is suggested that references to periodicals be furnished in some detail and in general in accordance with the standard adopted for the Subject Catalogue of the International Catalog of Scientific Literature, viz., name of author, with initials following (ordinarily omitting title of paper), abbreviated name of Journal, with place of publication, series (if any), volume, inclusive pages, year. For example: Montgomery, T. H., *J. Morph.*, Boston, **22**, 731–815 (1911); or, Wheeler, W. M., *Konigsberg, Schr. physik. Ges.*, **25**, 1–142 (1914).

MANUSCRIPTS will be accepted only from members of the Academy or from chairmen of the divisions of the National Research Council who will assume responsibility for the propriety and scientific standards of the paper and should be addressed to *Proceedings of the* NATIONAL ACADEMY OF SCIENCES, *2101 Constitution Ave., Washington, 25, D. C. Articles of which a member of the Academy is author or a co-author may extend to twelve printed pages, those by non-members only to six printed pages; space above these limits will be billed to the author at $17.50 per page.*

PROOF will ordinarily be sent,—with the understanding that charges for an author's correction shall be billed to him. Authors are therefore requested to make final revisions on the typewritten manuscripts.

REPRINTS should be ordered at the time of submission of manuscript. They will be furnished to authors at cost, approximately as follows:

SCHEDULE OF RATES FOR REPRINTS, CARRIAGE CHARGES EXTRA

No. Copies	50	100	150	200	250
4 pp.	4.90	5.60	6.30	7.00	7.70
8 pp.	9.80	11.20	12.60	14.00	15.40
Covers	3.75	4.80	5.85	6.90	7.95

CONTENTS

Page

BIOCHEMISTRY.—SEROLOGICAL REACTIVITY OF SYNTHETIC POLYGLUCOSES. . . . . *By Michael Heidelberger and Alan C. Aisenberg* 453

BIOCHEMISTRY.—A NEW PARTICLE TYPE IN CERTAIN CONNECTIVE TISSUE EXTRACTS . . . *By Francis O. Schmitt, Jerome Gross, and John H. Highberger* 459

CHEMISTRY.—HYDROGEN BONDED HELICAL CONFIGURATIONS OF THE POLYPEPTIDE CHAIN . . . . . *By Jerry Donohue* 470

GENETICS.—THE EFFECT OF A SUPPRESSOR ON ALLELIC INOSITOLLESS MUTANTS IN *Neurospora Crassa* . . . . . *By Norman H. Giles and C. W. H. Partridge* 479

GENETICS.—THE EFFECT OF PRETREATMENT WITH INFRA-RED RADIATION ON THE X-RAY INDUCED SEX-LINKED RECESSIVE LETHAL AND VISIBLE MUTATION RATE IN *Drosophila Melanogaster* . . . . . *By William J. Young, Henry T. Yost, Jr., P. T. Ives, and R. Paul Levine* 488

MATHEMATICS.—A THEORY OF *N*-PERSON GAMES WITH PERFECT INFORMATION . . . . . *By David Gale* 496

MATHEMATICS.—THE SPINOR CONNECTION IN THE UNIFIED EINSTEIN THEORY OF RELATIVITY . . . . . *By Václav Hlavatý* 501

MATHEMATICS.—CONNECTIONS BETWEEN EINSTEIN'S TWO UNIFIED THEORIES OF RELATIVITY . . . . . *By Václav Hlavatý* 507

MATHEMATICS.—ON THE CONTRACTION OF GROUPS AND THEIR REPRESENTATIONS . . . . . *By E. Inonu and E. P. Wigner* 510

MATHEMATICS.—ERGODIC PROPERTY OF THE BROWNIAN MOTION PROCESS. . . . . *By G. Kallianpur and H. Robbins* 525

MATHEMATICS.—OMNIBUS CHECKING OF THE 61-PLACE TABLE OF DENARY LOGARITHMS COMPILED BY PETERS AND STEIN, BY CALLET, AND BY PARKHURST. . . . . *By Horace S. Uhler* 533

MATHEMATICS.—SIGNIFICANCE LEVELS FOR A SKEW DISTRIBUTION. . . . . *By Edwin B. Wilson* 537

PHYSICS.—GENERALIZED RIEMANN SPACES AND GENERAL RELATIVITY . . . . . *By Luther P. Eisenhart* 546

PHYSICS.—A THEORY OF FERROMAGNETISM . . . . . *By Linus Pauling* 551

PHYSICS.—EQUAL RATIOS OF ENERGIES OF NUCLEAR EXCITED STATES OF EVEN-EVEN ISOTOPES . . . . . *By Martin G. Redlich* 560

PSYCHOLOGY.—THE EFFECT UPON HEART RATE CONDITIONING OF RANDOMLY VARYING THE INTERVAL BETWEEN CONDITIONED AND UNCONDITIONED STIMULI . . . . . *By P. J. Bersh, W. N. Schoenfeld, and J. M. Notterman* 565